



```

PF 22-OCT-1999; 99WO-US24046.
XX
XX 23-OCT-1998; 98US-0155209.
XX
XX (INCY-) INCYTE PHARM INC.
XX
XX Tang YT, Corley NC, Guegler KJ, Patterson C;
XX WPI: 2000-350699/30.
XX DR N-PSDB: AAA48963.
XX
XX Purified polypeptide used for treating or preventing a disorder
XX characterized by expression or activity of lysine-rich statherin
XX proteins -
XX
XX PS Claim 1; Page 69; 75pp; English.
XX
XX The present sequence is human lysine-rich statherin protein (LRSP). The
XX cDNA sequence encoding this protein was identified through analysis of
XX a cDNA library of breast tumor tissue (BRSTNOT14). The LRSP sequence
XX was found to have homology with human statherin (AA94527) and human
XX basic histidine-rich protein (AA94528). Human statherin is a
XX phosphoprotein that acts as an inhibitor of precipitation of calcium
XX phosphate salts in the oral cavity. The LRSP polypeptide and its
XX antagonists may be useful for treating or preventing disorders
XX associated with the activity of LRSP. Such disorders include
XX autoimmune/inflammatory disorders (for example AIDS, allergies, asthma,
XX diabetes mellitus), bacterial and fungal infection and cancers (such
XX as leukemia, adenocarcinoma, melanoma). Antibodies to LRSP may be
XX useful for diagnosis of the above disorders.
XX
XX SQ Sequence 95 AA;
XX
XX Query Match 100.0%; Score 522; DB 21; Length 95;
XX Best Local Similarity 100.0%; Pred. No. 1.8e-57;
XX Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 MFHFKVGRKQHFVETWETLSNNKTLVSLKKKPFHLYCYIYPLVPKLLIFLDIAFI 60
XX 1 MFHFKVGRKQHFVETWETLSNNKTLVSLKKKPFHLYCYIYPLVPKLLIFLDIAFI 60
XX
XX DB 1 MFHFKVGRKQHFVETWETLSNNKTLVSLKKKPFHLYCYIYPLVPKLLIFLDIAFI 60
XX
XX QY 61 PKSLISOFONNHYNHTNHTNNTNIRFNISNCR 95
XX 61 PKSLISOFONNHYNHTNHTNNTNIRFNISNCR 95
XX
XX DB 61 PKSLISOFONNHYNHTNHTNNTNIRFNISNCR 95
XX
XX RESULT 2
XX AAU25643
XX ID AAU25643 standard; Protein; 234 AA.
XX
XX AC AAU25643;
XX
XX DT 18-DEC-2001 (first entry)
XX
XX DE G protein-coupled receptor, ngPCR-2057.
XX
XX Human; mental disorder; thyroid disease; renal failure; anorexia;
XX inflammatory condition; Crohn's disease; rheumatoid arthritis; HIV;
XX autoimmune disorder; schizophrenia; migraine; stroke; dementia; obesity;
XX depression; Parkinson's disease; Alzheimer's disease; viral infection;
XX Huntington's disease; human immunodeficiency virus; type 2 diabetes;
XX anorexia; hypertension; thrombosis; myocardial infarction;
XX atherosclerosis; cancer; sexual dysfunction; G protein-coupled receptor;
XX ngPCR.
XX
XX OS Homo sapiens.
XX
XX PN WO200162924-A2.
XX
XX PD 30-AUG-2001.
XX
XX PF 23-FEB-2001; 2001WO-US05989.
XX

```

```

PR 24-FEB-2000; 2000US-0184602.
PR 24-FEB-2000; 2000US-0184604.
PR 24-FEB-2000; 2000US-0184606.
PR 24-FEB-2000; 2000US-0184609.
PR 24-FEB-2000; 2000US-0184650.
PR 24-FEB-2000; 2000US-0184710.
PR 24-FEB-2000; 2000US-0184712.
PR 24-FEB-2000; 2000US-0184715.
PR 24-FEB-2000; 2000US-0184716.
PR 24-FEB-2000; 2000US-0184725.
PR 24-FEB-2000; 2000US-0184822.
XX
XX (PHAA ) PHARMACIA & UPJOHN CO.
XX
XX PA
XX PI Vogell G, Wood LS, Parodi LA, Lind P;
XX
XX DR WPI: 2001-570632/64.
XX DR N-PSDB: AAS42953.
XX
XX PT Novel nucleic acid and encoded ngPCR-x, used to screen for compounds
XX for use in the treatment of mental disorders, such as Alzheimer's
XX disease, or Parkinson's disease -
XX
XX PS Claim 31; Page 69; 263pp; English.
XX
XX The invention relates to novel isolated human G protein-coupled
XX receptors (ngPCR-x). The ngPCR-x can be used for screening compounds
XX which can be used to treat mental disorders, thyroid disease, renal
XX failure, inflammatory conditions such as Crohn's disease, rheumatoid
XX arthritis, autoimmune disorders, schizophrenia, migraine, stroke,
XX dementia, depression, Parkinson's disease, Alzheimer's disease, and
XX Huntington's disease. They may also be used for treating viral infections
XX such as human immunodeficiency virus (HIV), type 2 diabetes, obesity,
XX anorexia, hypotension, hypertension, thrombosis, myocardial infarction,
XX atherosclerosis, cancer, and sexual dysfunction. AAU25617-AAU25726
XX represent the amino acid sequences of novel human G protein-coupled
XX receptors, ngPCR-2031 to ngPCR-2140 respectively, as described in the
XX invention.
XX
XX SQ Sequence 234 AA;
XX
XX Query Match 14.1%; Score 73.5; DB 22; Length 234;
XX Best Local Similarity 22.1%; Pred. No. 0.61;
XX Matches 21; Conservative 22; Mismatches 37; Indels 15; Gaps 3;
XX
XX QY 12 FKVTWETLSNNKTLVSLKKKPFHLYCYIYPLVPKLLIFLDIAFIKSLISQ 67
XX 12 FKVTWETLSNNKTLVSLKKKPFHLYCYIYPLVPKLLIFLDIAFIKSLISQ 67
XX
XX DB 133 FSNVLLKPGLSNSVSLNAEKRTITLIPSCVCIIFVLCILIPRSVFLSPHIKNCVSP 192
XX
XX QY 68 -----FONNHYNHTNHTNNTNIRFNIS 91
XX 68 -----FONNHYNHTNHTNNTNIRFNIS 91
XX
XX DB 193 LSLNPIMLWFKHQRHAIHAHGEPVOYCLIS 227
XX
XX RESULT 3.
XX AA37060
XX ID AA37060 standard; Protein; 310 AA.
XX
XX AC AA37060;
XX
XX DT 07-OCT-1999 (first entry)
XX
XX DE Chlamydia trachomatis lipoprotein sequence.
XX
XX Human; eye disease; conventional trachoma; nonendemic trachoma;
XX paratrachoma; inclusion conjunctivitis; genital disease; peritheatitis;
XX nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
XX bartolinitis; pneumonia; venereal lymphogranulomatosis.
XX
XX OS Chlamydia trachomatis.
XX
XX PN WO9928475-A2.
XX

```

[illegible]

DR WPI: 2002-179901/23.

PT Novel compositions comprising Chlamydia Cap1 protein and its use in the treatment of Chlamydia infection -

PS Claim 24; Page 322-323; 537pp; English.

XX The present invention describes compositions comprising a Chlamydia Cap1 protein and methods for the diagnosis and therapy of Chlamydia infection.

XX Chlamydia DNA and protein sequences from the present invention can have

XX antibacterial and immunostimulant activities, and can be used in

XX vaccines. Compounds from the present invention can be used for eliciting

XX an immune response, specifically stimulating a Chlamydia-specific T-cell

XX response or inhibiting the development of a Chlamydia infection in an

XX animal. Methods from the present invention can be used for detecting the

XX presence of Chlamydia in a patient; to stimulate and/or expand T cells

XX specific for a Chlamydia protein; and for treatment of a Chlamydia

XX infection. AB193394 to AB192709 and ABB94096 to ABB94374 represent

XX sequences used in the exemplification of the present invention.

XX

XX Sequence 412 AA;

XX

XX Query March 14.0%; Score 73; DB 23; Length 412;

XX Best Local Similarity 26.5%; Pred. NO. 1.4;

XX Matches 30; Conservative 19; Mismatches 34; Indels 30; Gaps 7;

XX

XX 2 WFH-KVGR-----KQHFKVTFWETDLSNKKTVLSLKKKPFHL-----YCVIYIPLV 47

XX 95 WEHKGAGRYSLLEKLSFRAS-----SSSEILIELEKPEPOLALIASFFFAVYRPN 148

XX

XX 48 PKLILFLFDIA-FIPKSLIS---OFONNHYTHNHTNNTNIRENITSNCRT 95

XX 149 P-----FLSSGPMPKTYVCGOTLVLOKNPYDYDHAVELHSIDFRILTPITYT 196

XX

XX RESULT 5

XX ABB94350

XX ID ABB94350 standard; Protein: 426 AA.

XX

XX ABB94350;

XX

XX 05-JUN-2002 (first entry)

XX

XX Chlamydia trachomatis protein sequence SEQ ID NO:573.

XX

XX Chlamydia infection; Chlamydia; vaccine; detection; diagnosis;

XX Chlamydia; antibacterial; immunostimulant; immune response;

XX Chlamydia-specific T-cell response.

XX

XX Chlamydia trachomatis.

XX

XX OS

XX PN W0200208267-A2.

XX

XX 31-JAN-2002.

XX

XX 20-JUL-2001; 2001WO-US23121.

XX

XX 20-JUL-2000; 2000US-0620412.

XX

XX 23-APR-2001; 2001US-0841132.

XX

XX (CORI-) CORIXA CORP.

XX

XX PA

XX PI Filing SP, Skeiky YAW, Probst P, Bhatia A;

XX

XX WPI: 2002-179901/23.

XX

XX Novel compositions comprising Chlamydia Cap1 protein and its use in the treatment of Chlamydia infection -

XX

XX Claim 34; Page 507-508; 537pp; English.

XX

XX The present invention describes compositions comprising a Chlamydia Cap1

CC Protein and methods for the diagnosis and therapy of Chlamydia infection.  
 CC Chlamydia DNA and protein sequences from the present invention can have  
 CC antibacterial and immunostimulant activities, and can be used in  
 CC vaccines. Compounds from the present invention can be used for eliciting  
 CC an immune response, specifically stimulating a Chlamydia-specific T-cell  
 CC response or inhibiting the development of a Chlamydia infection in an  
 CC animal. Methods from the present invention can be used for detecting the  
 CC presence of Chlamydia in a patient; to stimulate and/or expand T cells  
 CC specific for a Chlamydia protein; and for treatment of a Chlamydia  
 CC infection. ABL92394 to ABL92709 and ABB94096 to ABB94374 represent  
 CC sequences used in the exemplification of the present invention.  
 CC XX

SO Sequence 426 AA;

Query Match 14.0%; Score 73; DB 23; Length 426;  
 Best Local Similarity 26.5%; Pred. No. 1.4;  
 Matches 30; Conservative 19; Mismatches 34; Indels 30; Gaps 7;

OY 2 WEH--KVGR-----KQHKVTFWETDLSNNKTVSLKKKPFHL-----YCVIYIPLV 47  
 DB 109 WEHTKQAGRYSLLEKLSFRAS-----SSSEILLEKEPEQLALIASPFVAYRREN 162  
 OY 48 PKLILFLDLA-FIPKSLIS----OFQNNHYTHNHTNNTNINRFINISNCR 95  
 DB 163 P-----FLSSGPFMPKTYVGQTLVLOKNPYYDHAEHLSDIRIIPNIYT 210

RESULT 6

ABB94255 ID ABB94255 standard; Protein; 433 AA.

AC ABB94255;

DT 05-JUN-2002 (first entry)

DE Chlamydia protein sequence SEQ ID NO:363.

KW Chlamydia infection; Chlamydia; vaccine; detection; diagnosis;  
 KW antigen; antibacterial; immunostimulant; immune response;

KW Chlamydia-specific T-cell response.

OS Chlamydia sp.

PN W0200208267-A2.

PD 31-JAN-2002.

PE 20-JUL-2001; 2001WO-US23121.

PR 20-JUL-2000; 2000US-0620412.

PR 23-APR-2001; 2001US-0841132.

PA (CORI-) CORIXA CORP.

PI Filing SP, Skelky YAW, Probst P, Bhatia A;

DR WPI; 2002-179901/23.

PT Novel compositions comprising Chlamydia Cap1 protein and its use in the  
 PT treatment of Chlamydia infection -

PS Claim 24; Page 323-324; 537pp; English.

CC The present invention describes compositions comprising a Chlamydia Cap1  
 CC protein and methods for the diagnosis and therapy of Chlamydia infection.  
 CC Chlamydia DNA and protein sequences from the present invention can have  
 CC antibacterial and immunostimulant activities, and can be used in  
 CC vaccines. Compounds from the present invention can be used for eliciting  
 CC an immune response, specifically stimulating a Chlamydia-specific T-cell  
 CC response or inhibiting the development of a Chlamydia infection in an  
 CC animal. Methods from the present invention can be used for detecting the  
 CC presence of Chlamydia in a patient; to stimulate and/or expand T cells  
 CC specific for a Chlamydia protein; and for treatment of a Chlamydia

CC infection. ABL92394 to ABL92709 and ABB94096 to ABB94374 represent  
 CC sequences used in the exemplification of the present invention.  
 CC XX

SO Sequence 433 AA;

Query Match 14.0%; Score 73; DB 23; Length 433;  
 Best Local Similarity 26.5%; Pred. No. 1.5;  
 Matches 30; Conservative 19; Mismatches 34; Indels 30; Gaps 7;

OY 2 WEH--KVGR-----KQHKVTFWETDLSNNKTVSLKKKPFHL-----YCVIYIPLV 47  
 DB 116 WEHTKQAGRYSLLEKLSFRAS-----SSSEILLEKEPEQLALIASPFVAYRREN 169  
 OY 48 PKLILFLDLA-FIPKSLIS----OFQNNHYTHNHTNNTNINRFINISNCR 95  
 DB 170 P-----FLSSGPFMPKTYVGQTLVLOKNPYYDHAEHLSDIRIIPNIYT 217

RESULT 7

AAW13825 ID AAW13825 standard; Protein; 1226 AA.

AC AAW13825;

DT 04-JUN-1997 (first entry)

DE Yeast transcription regulatory factor SRB8.

KW Transcription regulatory factor; suppressor of RNA polymerase B;  
 KW SRB8; RNA polymerase II; holoenzyme; SWI/SNF.

OS Saccharomyces cerevisiae.

PN W09708301-A1.

PD 06-MAR-1997.

PE 28-AUG-1996; 96WO-US14192.

PR 26-JAN-1996; 96US-0590399.

PR 31-AUG-1995; 95US-0521872.

PR 11-OCT-1995; 95US-0540804.

PA (WHED) WHITEHEAD INST BIOMEDICAL RES.

PI Chao DM, Koleske AJ, Thompson CM, Young RA;

DR WPI; 1997-179258/16.

DR N-PDB; AAT59908.

PT Purified RNA polymerase II holoenzyme - comprises RNA polymerase II  
 PT and one or more regulatory proteins, pref. suppressor of RNA  
 PT polymerase B proteins or SWI/SNF proteins

PS Claim 11; Fig 10a-b; 154pp; English.

CC Novel yeast SRB (suppressor of RNA polymerase B) proteins SRB4, SRB5,  
 CC SRB6, SRB7, SRB8, SRB9, SRB10 and SRB11 (AAW13821-28) are transcription  
 CC regulatory factors that act as positive and negative regulators of  
 CC RNA polymerase II activity, and are components of the RNA polymerase  
 CC II holoenzyme. They were identified using methods designed to  
 CC identify transcription factors involved in RNA polymerase II  
 CC C-terminal domain (CTD) function. SRB8 and SRB9 appear to repress  
 CC CTD activity. Genomic clones (AAT59904-11) for the SRBs have been obtd.  
 CC SRBs can be used to treat diseases resulting from alteration or  
 CC deletion of the SRB gene, pref. by gene transfer technology. They  
 CC can also be used in in vitro transcription of DNA and to identify  
 CC cpds. that modify gene transcription.  
 CC XX

SO Sequence 1226 AA;

Query Match 13.6%; Score 71; DB 18; Length 1226;  
 Best Local Similarity 27.9%; Pred. No. 8.9;



Matches	24;	Conservative	14;	Mismatches	32;	Indels	16;	Gaps	4;
Qy	11	HEVVTW-----	EDLSNNKTVLSIKKKRPHLYGVYIPLVPKLIILFDIAFTPKSL	64					
Db	82	HLIMFMIMDQCIDITNAPVAARITTS-SQKEPFP-----	LVTKITDMLLHKYIVSSS	132					
Qy	65	ISOFONNHYTHNHTNHTNINRENT	90						
Db	133	KSKINDENYIINHKKV-NKIKLNL	157						
RESULT 8									
ABB09493									
ID	ABB09493	standard; Protein; 1381 AA.							
XX	ABB09493;								
XX	01-JUL-2002	(first entry)							
XX	AMEPV	ABC transporter-like protein (AMV130) amino acid sequence.							
XX	DE								
XX	AMEPV; gene therapy; viral vector; chromosome mapping; gene mapping;								
XX	KW	genetic deficiency disorder; ABC transporter-like protein.							
XX	OS	Amsacta moorei entomopoxvirus.							
XX	Key	Location/Qualifiers							
XX	FT	Misc-difference 268..269							
XX	FT	/note="encoded by TTGTGTAAT"							
XX	FT	Misc-difference 337..338							
XX	FT	/note="encoded by AATATTAA"							
XX	FT	Misc-difference 476..477							
XX	FT	/note="encoded by AATAATTG"							
XX	FN	W0200212526-A2.							
XX	PD	14-FEB-2002.							
XX	PF	10-AUG-2001; 2001WO-US25287.							
XX	PR	10-AUG-2000; 2000US-224479P.							
XX	PR	14-SEP-2000; 2000US-0662254.							
XX	PA	(UFL) UNIV FLORIDA.							
XX	PI	Moyer RW, Li Y, Bawden AL;							
XX	DR	WPI; 2002-227161/28.							
XX	DR	N-PSDB; ABL55640.							
XX	PT	Novel recombinant entomopox virus vector useful for delivering							
XX	PT	polynucleotide encoding protein to vertebrate cell, comprises							
XX	PT	polynucleotide encoding protein operably linked with heterologous							
XX	PT	promoter sequence							
XX	PS	Claim 71; Page 118-122; 326pp; English.							

Query Match	13.3%	Score 69.5	DB 23	Length 1381
Best Local Similarity	32.9%	Pred. NO. 16		
Matches 23	Conservative 8	Mismatches 26	Indels 13	Gaps 2
Qy	30 LKKKKPFHLYCYVYIPVYKLTLLFLDIAFIKPSLISOPONNHYTHNHTN-----	83		
Db	16 IRRKK-----YIDMLEYLLILLFLFTLLYSKKNKIKYDNDLNNINNTNTNTIITYPK 69			
Qy	84 -NIRENIIISN 92			
Db	70 SNISIKIIEEN 79			
RESULT 9				
AAB18244	ID AAB18244 standard; Protein; 1245 AA.			
XX	AAB18244;			
XX	07-NOV-2000 (first entry)			
XX	Plasmodium falciparum, chromosome 2 related protein SEQ ID NO:101.			
XX	Plasmodium falciparum, chromosome 2; human malaria parasite; vaccine;			
KW	antimalarial; malaria; protozoacide; infection; insecticide.			
XX	Plasmodium falciparum.			
OS	WO200025728-A2.			
XX	11-MAY-2000.			
XX	05-NOV-1999; 99WO-US26796.			
XX	05-NOV-1998; 98US-0107131.			
PA	(HOEF/) HOFFMAN S.			
PA	(CARU/) CARUCCI D.			
PA	(GARD/) GARDNER M.			
PA	(VENT/) VENTER J C.			
PI	Hoffman S, Carucci D, Gardner M, Venter JC;			
XX	WPI: 2000-365347/31.			
XX	Proteins encoded by chromosome 2 of the human malarial parasite,			
XX	Plasmodium falciparum, useful as antimalarial vaccines and in the			
XX	diagnosis of P.falciparum infection -			
XX	Disclosure; Page 234-237; 577Pp; English.			
XX	The present invention describes proteins and their fragments (I) encoded			
XX	by chromosome 2 of the human malarial parasite, Plasmodium falciparum.			
XX	Also described are: (I) nucleotide sequences (II) encoding (I); and (2)			
XX	vaccines against P. falciparum infection comprising (I) or (II).			
XX	(I) and (II) are useful for the development of vaccines against			
XX	P. falciparum infection. (I) and polyclonal antisera or a monoclonal			
XX	antibody raised to immunogens comprising the sequences of (I), are			
XX	useful in the detection of infection with P. falciparum. Furthermore,			
XX	(I) (especially when they are rifins or secreted or membrane proteins)			
XX	can aid the identification of drugs to treat or prevent P. falciparum			
XX	infection, or they can be used to identify drug resistance in			
XX	P. falciparum. Sequencing of the plasmodium chromosome 2 and the			
XX	subsequent identification of proteins encoded by it will help to expand			
XX	our understanding of parasite biology, a process hampered by the			

CC complexity of the parasitic lifecycle, and provide new targets for  
 CC vaccine and drug development. Parasite resistance to drugs and mosquito  
 CC resistance to insecticides have led to a resurgence of malaria in many  
 CC parts of the world, and there is a pressing need for vaccines and new  
 CC drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide  
 CC and protein sequences given in the present invention, but which are not  
 CC specifically mentioned within the specification.

SO Sequence 1245 AA;

Query Match 13.2%; Score 69; DB 21; Length 1245;  
 Best Local Similarity 27.8%; Pred. No. 16;

Matches 25; Conservative 12; Mismatches 33; Indels 20; Gaps 3;

OY 22 SNNKTLVSLKKRPFHLYCIVYIPVPLKLIILFDLAFKSLSCFQNNHYTHN--T 78  
 DB 1069 SNNKFTIIRKKKKLKKLCLCYIMKSEPHI---LDEFWMLSCONEIKNIYKLNLEHVIS 1124  
 OY 79 NHTNNTNIRFNIIS-----NCRT 95  
 DB 1125 LHNSSIIDFKIINHFLNKIFENISINCT 1154

RESULT 10

AAW98706 AAW98706 standard; Protein; 346 AA.

AC AAW98706;

DT 31-MAR-1999 (first entry)

DE H. pylori GHP0 727 protein.

KW GHP0 protein; Helicobacter infection; gastroduodenal disease; gastritis;  
 KM peptic ulcer disease.

OS Helicobacter pylori.

PN WO9843478-A1.

PD 08-OCT-1998.

PF 01-APR-1998; 98WO-US06371.

PR 29-JUL-1997; 97US-0902615.

PR 01-APR-1997; 97US-0833457.

PR 24-JUN-1997; 97US-0881227.

PA (HUMA-) HUMAN GENOME SCI INC.

XX (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.

PI Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;

DR WPI: 1998-542293/46.

DR N-PSDB; AAX14425.

PT New isolated Helicobacter polynucleotides - used to develop products  
 PT for the diagnosis, prevention and treatment of Helicobacter  
 PT infections and gastrointestinal diseases

PS Claim 8; Page 1482-1484; 2054pp; English.

CC This sequence represents a Helicobacter pylori GHP0 protein of the  
 CC invention. The polypeptides can be used for preventing or treating  
 CC Helicobacter infections, and gastroduodenal diseases associated with  
 CC these infections, including acute, chronic, and atrophic gastritis, and  
 CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be  
 CC used for the production of antibodies. The products can also be used for  
 CC detection and diagnosis.

XX Sequence 346 AA;

Query Match 13.0%; Score 68; DB 19; Length 346;

Best Local Similarity 26.3%; Pred. No. 4.7;  
 Matches 26; Conservative 12; Mismatches 45; Indels 16; Gaps 3;

OY 11 HFVY-----TFWETDLSNNKTLVSLKKRPFHLYCIVYIPVPLKLIIL-----FLDIAR 59  
 DB 146 HFVKSIVSTYVAVSGAGNKGJESLKNKLTALCELEKPTIDLNQVLAQAFAYPIAR 205

OY 60 IPKSLISQFQNNHYTHN-----HTNNTNIRFNIISNC 93

DB 206 NAIAMHIDTFRENGVTKELKMLHETKIMGVDFPISATC 244

RESULT 11

ABBA9049 ABB49049 standard; Protein; 907 AA.

AC ABB49049;

DT 05-FEB-2002 (first entry)

DE Listeria monocytogenes protein #1753.

KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;  
 KM vitamin B12; bacterial infection; disease.

OS Listeria monocytogenes.

PN WO200177335-A2.

PD 18-OCT-2001.

PF 11-APR-2001; 2001WO-FR01118.

PR 11-APR-2000; 2000FR-0004629.

PA (INSP) INST PASTEUR.

PI Buchrieser C, Frangeul L, Couve E, Rusnok C, Fsihi H, Deloux P;  
 PI Dussurget O, Chetoui A, Nedjati H, Glaser P, Kunst F, Cossart P;  
 PI Daniels J, Goebel W, Kieft M, Kuhn M, Ng E, Vazquez-Boland JA;  
 PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;  
 PI Chakraborty T, Doman E, Hain T, Berche P, Charbit A, Durant L;  
 PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;  
 PI Madueno E, De Pablo B, Wehlund J, Kaerst U, Entian K, Hauf J;  
 PI Rose M, Voss H;

DR WPI: 2002-010914/01.

PT Genomic sequence for Listeria monocytogenes, useful e.g. for treatment  
 PT and prevention of Listeria and related bacterial infections, and  
 PT related polypeptides

PS Claim 6; SEQ ID No 1754; 192pp; French.

CC The present invention relates to the genome sequence of Listeria  
 CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of  
 CC it are useful for selecting probes and primers for detecting genes in L.  
 CC monocytogenes and related organisms, and for studying genetic  
 CC polymorphisms and other genomes. The present sequence is a protein  
 CC encoded by the genome sequence of the present invention. Proteins  
 CC expressed from the genome sequence are useful for raising specific  
 CC antibodies, identification of L. monocytogenes and related organisms,  
 CC for biosynthesis and biodegradation, especially biosynthesis of vitamin  
 CC B12. The genome sequence and proteins encoded by it are also useful for  
 CC selecting compounds that regulate gene expression and cell replication  
 CC and modulate L. monocytogenes-related diseases. In addition, the genome  
 CC sequence and proteins encoded by it are useful in pharmaceutical and  
 CC vaccines compositions for the treatment or prevention of infections by L.  
 CC monocytogenes and related organisms.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pat\_sequences.

## RESULT 13

XW		cell typing; abnormal cell function; ss.
XX		Homo sapiens.
XS		
XP		WO9514772-1.
PN		
PD		01-JUN-1995.
PP		
PE		11-NOV-1994; 94MO-JP01916.
PF		12-NOV-1993; 93JP-0355504.
PR		(MATS/) MATSUBARA K.
PS		(OKUBO/) OKUBO K.
PT		Matsubara K, Okubo K;
PI		WPI: 1995-206931/27.
PJ		
PK		Identifying gene signatures in 3'-directed human cDNA library - e.g
PL		for diagnosis of abnormal cell function, by preparing cDNA that
PM		reflects relative abundance of corresp. mRNA in specific human
PN		tissues
PO		
PP		Claim 1; Page 1582; 2245pp; Japanese.
PS		
PT		A single-stranded DNA (or its complementary strand or the corresp.
PU		double-stranded DNA) which comprises one of the 7837 "GS" sequences
PV		given in AAT19001-T26837 and which is able to hybridize to part of
PX		human genomic DNA, cDNA or mRNA is claimed. The GS (Gene signature)
PY		sequences were obtained from 3'-directed cDNA libraries prepared
PZ		from various human tissues; synthesis of cDNA was initiated from the
QA		3'-end of mRNA by using poly(T) as the sole primer. Since the 3'
QB		untranslated sequence is unique to a particular mRNA species, almost
QC		all the 3'-oriented cDNAs hybridize with specific mRNAs. Each libra-
QD		ry is constructed so as to reflect accurately the relative abundance of
QE		different mRNAs in the particular tissue from which it was derived.
QF		The appearance frequency of a given GS in a cDNA library can be
QG		determined (esp. using primers and probes derived from the GS
QH		sequences) as a means of diagnosing abnormal cell function or for
QI		recognising different cell types.
QJ		
QK		Sequence 126 BP; 44 A; 16 C; 25 G; 34 T; 7 other;
QL		
QM		Alignment Scores:
QN		Pred. No.: 7.9e-06 Length: 126
QO		Score: 117.00 Matches: 20
QP		Percent Similarity: 78.5% Conservative: 2
QQ		Best Local Similarity: 71.4% Mismatches: 6
QR		Query Match: 22.4% Indels: 0
QS		DB: 16 Gaps: 0
QT		
QU		US-09-830-244B-1 (1-95) x AAT24320 (1-126)
QV		
QW		1 MetTPhheHisysyAGLyArgylsglnHisPhelysValThrphetpGlnrrasp 20
QX		.....:       .....:     ..
QY		26 ATGGGGCTTCACCAACTGGGAAGAAGCAGCAATTAAATANCTTTTGCGNGACTGAN 85
QZ		.....:       .....:     ..
RA		Db 21 LeusEsrAsnsllysthrleuVal 28
RB		.....:       .....:     ..
RC		86 TTGGGTATATATATTAANTCTTCAGTC 109
RD		
RE		RESULT 8
RF		ABL54353
RG		ABL54353 standard; DNA; 9881 BP.
RH		XX ABL54353;
RI		XX
RJ		DT 29-JUL-2002 (first entry)
RK		XX
RL		DE Chemically treated apoptosis gene #27.
RM		XX
RN		Apoptosis; HIV; Bloom syndrome; Cardioathv.

XX WO200234771-A2.  
 XX 02-MAY-2002.  
 PD 29-OCT-2001; 2001WO-GB04789.  
 PF 27-OCT-2000; 2000GB-0026333.  
 PR 24-NOV-2000; 2000GB-0028727.  
 PR 07-MAR-2001; 2001GB-0005640.  
 XX (CHIR-) CHIRON SPA.  
 PA (GENO-) INSR GENOMIC RES.  
 XX Telford J, Maignan V, Margalit Ros YI, Grandi G, Fraser C;  
 PI Tettelein H;  
 XX WPI; 2002-352536/38.  
 DR N-PSDB; ABN68051.  
 XX New Streptococcus protein for the treatment or prevention of infection  
 PT or disease caused by Streptococcus bacteria, such as meningitis, and  
 PT for detecting a compound that binds to the protein -  
 XX Claim 1; Page 3557; 4525pp; English.  
 PS  
 XX The invention relates to a protein (ABP25413-ABP30895) from group B  
 CC Streptococcus/GAS (Streptococcus agalactiae) or group A Streptococcus/GAS  
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
 CC the specification. The proteins have antibacterial and anti-inflammatory  
 CC activity. (1), nucleic acids encoding (1), ABN66044-ABN71536 and  
 CC antibodies that bind (1) are used in the manufacture of medicaments for  
 CC the treatment or prevention of infection or disease caused by  
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
 CC Nucleic acids encoding (1) are used to detect Streptococcus in a  
 CC biological sample. (1) is used to determine whether a compound binds to  
 CC (1). A composition comprising (1) or a nucleic acid encoding (1), may be  
 CC used as a vaccine or diagnostic composition. The disease caused by  
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
 CC acid encoding (1) may be used to recombinantly produce (1) and may be  
 CC used in gene therapy. Antibodies to (1) are used for affinity  
 CC chromatography, immunoassays, and distinguishing/identifying  
 CC Streptococcus proteins.  
 CC  
 XX Sequence 244 AA;  
 SO  
 Query Match 12.6%; Score 66; DB 23; Length 244;  
 Best Local Similarity 26.0%; Pred. No. 5.5;  
 Matches 34; Conservative 14; Mismatches 41; Indels 42; Gaps 8;  
 OY 2 WFHVGV---RKQHF-----KVTWETDLSNNK---TLVSLKKK 34  
 DB 107 WLTIGESKRKTHFELLCALHELVKNSMPPIVILTTDFINNETLKDYILQSLSTDK 166  
 OY 35 -PEHLVCVI-----YIPVPLKLTIT---LELDIAFIPKSLISOFONNHTHTNHNNTN 83  
 DB 167 IREHSYLLDNTISQISTLKPDLITNRKLF---TYTTELATNSLVAHIDVDTSPQS 223  
 OY 84 NIRENITSNCR 94  
 DB 224 NTO-TTISNIK 233  
 RESULT 15  
 ABP07000  
 ID ABP07000 standard; Protein; 61 AA.  
 XX AC ABP07000;  
 XX DT 24-JUN-2002 (first entry)  
 XX DE Human ORFX protein sequence SEQ ID NO:13982.  
 XX

KW Human: open reading frame; ORFX: gene therapy; cancer; cirrhosis;  
 KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;  
 KW degenerative disorder; osteoarthritis; neurodegenerative disorder;  
 KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;  
 KW hypertension; hypothyroidism; cholesterol ester storage disease;  
 KW immune deficiency; immune disorder; infectious disease;  
 KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;  
 KW myasthenia gravis.  
 OS Homo sapiens.  
 XX WO200192523-A2.  
 XX 06-DEC-2001.  
 PD 29-MAY-2001; 2001WO-US10836.  
 PF 30-MAY-2000; 2000US-206132P.  
 PR 29-AUG-2000; 2000US-228716P.  
 XX (CURA-) CURAGEN CORP.  
 PA Shinkets RA, Leach MD;  
 PI WPI; 2002-106308/14.  
 DR N-PSDB; ABN22752.  
 XX Novel human polypeptides and polynucleotides useful for diagnosing,  
 PT preventing and treating cardiovascular disease, neurodegenerative,  
 PT hyperproliferative disorders and autoimmune disorders -  
 XX Disclosure: SEQ ID 13982; 1037pp; English.  
 PS  
 XX The present invention describes substantially purified human proteins  
 CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1  
 CC in the specification). ABN15762 to ABN27252 encode the human ORFX  
 CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for  
 CC treating or preventing a pathology associated with an ORFX-associated  
 CC disorder in humans, and in the manufacture of a medicament for treating a  
 CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide  
 CC sequences can be used in gene therapy. ORFX sequences can be used in the  
 CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,  
 CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,  
 CC osteoarthritis, neurodegenerative disorders, disorders related to organ  
 CC transplantation, cardiovascular diseases, diabetes mellitus, systemic  
 CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester  
 CC storage disease, various immune deficiencies and disorders, infectious  
 CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid  
 CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host  
 CC disease and autoimmune inflammatory eye disease. ORFX proteins are also  
 CC useful for treating burns, incisions, ulcers, for treating osteoporosis,  
 CC bone degenerative disorders, or periodontal disease, and for gut  
 CC protection or regeneration in various tissues and conditions resulting from  
 CC reperfusion injury in various tissues and conditions.  
 CC N.B. The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WPI  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX Sequence 61 AA;  
 SO  
 Query Match 12.5%; Score 65.5; DB 23; Length 61;  
 Best Local Similarity 34.0%; Pred. No. 1.2;  
 Matches 16; Conservative 7; Mismatches 23; Indels 1; Gaps 1;  
 OY 37 HLXCVIIPVPLKLTIT---LELDIAFIPKSLISOFONNHTHTNHNNTN 83  
 DB 3 NLXCVTLGIKPSVSYTLVGAQAIKPSK-DPLQOHQTHNATQSLNS 48  
 Search completed: April 14, 2003, 22:59:49  
 Job time: 111 secs

Tue Apr 15 15:15:18 2003

us-09-830-244b-1.ra1

Page. 1

GenCore version 5.1.4-p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2003, 23:01:31 ; Search time 90 Seconds  
(without alignments)  
31.058 Million cell updates/sec

Title: US-09-830-244B-1  
Perfect score: 522  
Sequence: 1 MWFHKYGRKQHFVFEWED.....NHTNHTNIRNITSNCT 95

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues  
Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents-AA:\*  
1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PTDUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/Backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	65	12.5	402	2	US-08-290-731C-14
2	65	12.5	911	4	US-09-356-952-6
3	64.5	12.4	445	4	US-09-134-001C-3507
4	63.5	12.2	2787	4	US-09-245-041-15
5	61.5	11.8	863	4	US-09-619-353-14
6	61	11.7	110	2	US-08-569-166-34
7	61	11.7	1226	2	US-08-540-804-12
8	61	11.7	1226	3	US-08-218-265-12
9	61	11.7	1226	3	US-08-521-872-12
10	61	11.7	1226	4	US-08-590-399-12
11	60.5	11.6	166	4	US-09-615-192A-273
12	60.5	11.6	332	2	US-08-671-978A-8
13	60	11.5	429	4	US-08-795-876-44
14	60	11.5	474	1	US-08-453-742-27
15	60	11.5	474	1	US-08-453-742-27
16	60	11.5	474	1	US-08-453-742-27
17	60	11.5	474	1	US-08-453-742-27
18	60	11.5	474	1	US-08-453-742-27
19	60	11.5	477	1	US-08-454-464-27
20	60	11.5	477	1	US-08-453-742-27
21	60	11.5	477	1	US-08-453-742-27
22	59.5	11.4	485	1	US-07-991-867B-42
23	59.5	11.4	485	1	US-07-991-867B-42
24	59.5	11.4	485	1	US-08-544-332-42
25	59.5	11.4	485	1	US-09-370-861A-42
26	59.5	11.4	485	1	US-09-370-861A-42
27	59	11.3	99	2	US-08-422-333-3

28	59	11.3	695	1	US-08-339-152A-30	Sequence 30, Appl
29	59	11.3	695	4	US-09-458-481B-4	Sequence 4, Appl
30	59	11.3	695	4	US-09-458-481B-5	Sequence 5, Appl
31	59	11.3	695	4	US-09-458-481B-6	Sequence 6, Appl
32	59	11.3	1311	3	US-08-340-011-5	Sequence 3, Appl
33	59	11.3	1311	3	US-08-901-710-5	Sequence 5, Appl
34	59	11.3	2391	2	US-08-446-855A-2	Sequence 2, Appl
35	59	11.3	2391	4	US-09-150-741-2	Sequence 2, Appl
36	59	11.3	3335	1	US-08-323-170B-2	Sequence 2, Appl
37	59	11.3	3335	1	US-08-954-441-2	Sequence 2, Appl
38	58.5	11.2	281	4	US-09-134-001C-4763	Sequence 4763, Ap
39	58.5	11.2	484	4	US-09-134-001C-5063	Sequence 5063, Ap
40	58.5	11.2	735	2	US-08-765-243-6	Sequence 6, Appl
41	58.5	11.2	735	5	PCT-US95-07285-6	Sequence 9, Appl
42	58.5	11.2	921	4	US-09-206-800-9	Sequence 1, Appl
43	58.5	11.2	1164	2	US-08-589-756-1	Sequence 1, Appl
44	58.5	11.2	1164	4	US-09-206-800-1	Sequence 1, Appl
45	58.5	11.2	1164	4	US-09-206-898-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1  
US-08-290-731C-14  
Sequence 14, Application US/08290731C  
Patent No. 5843646  
GENERAL INFORMATION:  
APPLICANT: BOWTELL, David Douglas Lawrence  
TITLE OF INVENTION: DNA MOLECULES ENCODING MORINE  
TITLE OF INVENTION: SON OF SEVENLESS (MSOS) GENE,  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESS: SUGHRUE, MION, ZINN, MACPEAK & SEAS  
STREET: 2100 PENNSYLVANIA AVENUE, N.W.  
CITY: WASHINGTON  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/290,731C  
FILING DATE: 17-OCT-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/AU93/00068  
FILING DATE: 17-FEB-1993  
APPLICATION NUMBER: PCT/US92/192  
FILING DATE: 17-FEB-1992  
AUTHOR/AGENT INFORMATION:  
NAME: KIT, Gordon  
REGISTRATION NUMBER: 30,764  
REFERENCE/DOCKET NUMBER: Q-36066  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 293-7060  
TELEFAX: (202) 293-7860  
TELEX: 6491103  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 402 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-290-731C-14  
Query Match 12.5% Score 65; DB 2; Length 402;  
Best Local Similarity 25.2% Pred. No. 3.8;

Matches 28; Conservative 10; Mismatches 23; Indels 50; Gaps 5;

QY 23 NNTLVSL-----KKKPPHLY-----CVIYIPVPLKILFLDLAIFPK-----62  
 Db 284 NSKTLASFELNNLTARKNFSYRDCLENCY--LPCVPLGVYFTDLFLKTKGKDNFQ 341

QY 63 -----SLISQFONNHYTHNHTN-----HNTNNI 85  
 Db 342 NMINEFKRTVTRILNEIKKFQSYGMFNPINEVOELLNEVISRENTNNI 392

RESULT 2  
 US-09-356-952-6  
 ; Sequence 6, Application US/09356952  
 ; Patent No. 6117663  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Borjacks-Sjodin, Ann  
 ; APPLICANT: Margaret, S. M.  
 ; APPLICANT: Bor-Sogi, Dafna  
 ; APPLICANT: Cole, Philip  
 ; APPLICANT: Kurliyan, John  
 ; TITLE OF INVENTION: A CRYSTAL OF A RAS-SOS COMPLEX AND METHODS OF USE  
 ; FILE REFERENCE: 600-1-228N  
 ; CURRENT FILING DATE: 1999-07-19  
 ; EARLIER FILING DATE: 1998-07-21  
 ; NUMBER OF SEQ ID NOS: 14  
 ; SOFTWARE: Patentln Ver. 2.0  
 ; SEQ ID NO 6  
 ; LENGTH: 911  
 ; TYPE: PRT  
 ; ORGANISM: Schizosaccharomyces pombe  
 US-09-356-952-6

Query Match Best Local Similarity 12.5%; Score 65; DB 3; Length 911;  
 Matches 28; Conservative 10; Mismatches 23; Indels 50; Gaps 5;

QY 23 NNTLVSL-----KKKPPHLY-----CVIYIPVPLKILFLDLAIFPK-----62  
 Db 776 NSKTLASFELNNLTARKNFSYRDCLENCY--LPCVPLGVYFTDLFLKTKGKDNFQ 833

QY 63 -----SLISQFONNHYTHNHTN-----HNTNNI 85  
 Db 834 NMINEFKRTVTRILNEIKKFQSYGMFNPINEVOELLNEVISRENTNNI 884

RESULT 3  
 US-09-134-001C-3507  
 ; Sequence 3507, Application US/09134001C  
 ; Patent No. 6380370  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lynn Doucette-Stamm et al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
 ; FILE REFERENCE: GTC-007  
 ; CURRENT FILING DATE: 1998-08-13  
 ; PRIOR FILING DATE: 1997-11-08  
 ; PRIOR APPLICATION NUMBER: US 60/064,964  
 ; PRIOR FILING DATE: 1997-08-14  
 ; NUMBER OF SEQ ID NOS: 5674  
 ; SEQ ID NO 3507  
 ; LENGTH: 445  
 ; TYPE: PRT  
 ; ORGANISM: Staphylococcus epidermidis  
 US-09-134-001C-3507

Query Match Best Local Similarity 12.4%; Score 64.5; DB 4; Length 445;  
 Matches 25.0%; Pred. No. 5;

Matches 21; Conservative 18; Mismatches 26; Indels 19; Gaps 4;

QY 12 FKVTMETDIS-----NNTLVSLKKKPPHLYCVIYIPVPLKILFLDLAIFPKSLI 65  
 Db 168 YKISYMLTNAVAFGALLGSLWGAHKSMLFFIAFYIYIMVFLALIV-----WLPKDL- 220

QY 66 SFQNNHYTHNHTNNTNNIRENI 89  
 Db 221 -----NIVQSHI--HNAEKQFSM 238

RESULT 4  
 US-09-245-041-15  
 ; Sequence 15, Application US/09245041  
 ; Patent No. 6274339  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Moore, K.  
 ; APPLICANT: Nagle, D.  
 ; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE DIAGNOSIS AND TREATMENT  
 ; FILE REFERENCE: 7853-136  
 ; CURRENT FILING DATE: 1999-02-05  
 ; EARLIER FILING DATE: 1998-07-21  
 ; EARLIER FILING DATE: 1998-07-21  
 ; NUMBER OF SEQ ID NOS: 131  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 15  
 ; LENGTH: 2787  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-09-245-041-15

Query Match Best Local Similarity 12.2%; Score 63.5; DB 4; Length 2787;  
 Matches 29; Conservative 11; Mismatches 25; Indels 31; Gaps 6;

QY 5 KVRKQHPKV-----TFWETDL-----SN--NKTLSLKKKPPHLYCVIYIPVPLKIL 53  
 Db 1611 LKKGKGYKCGQGCYFECKLVVVFYSNLNLVLIHLKLSIFSL-----FVPLHL 1663

QY 54 FLDAIFPKSLISQFONNHYTHNHTNNTNNIRENI 89  
 Db 1664 TLPUSHI-----HTQHT--HNTKCLDFKI 1686

RESULT 5  
 US-09-619-353-14  
 ; Sequence 14, Application US/09619353  
 ; Patent No. 6410249  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ngai, John  
 ; APPLICANT: Speca, David J.  
 ; APPLICANT: Lin, David M.  
 ; APPLICANT: Isaacoff, Ehud Y.  
 ; APPLICANT: Dittman, Andrew H.  
 ; APPLICANT: Fan, Jinhong  
 ; TITLE OF INVENTION: Ocular Receptors  
 ; FILE REFERENCE: B99-038-2  
 ; CURRENT FILING DATE: 2000-07-19  
 ; PRIOR FILING DATE: 1999-07-20  
 ; NUMBER OF SEQ ID NOS: 14  
 ; SOFTWARE: Patentln Ver. 2.1  
 ; SEQ ID NO 14  
 ; LENGTH: 863  
 ; TYPE: PRT  
 ; ORGANISM: Brachydanio rerio (zebrafish)  
 US-09-619-353-14



; sequence 12, Application US/08540804

STREET: Two Militia Drive





TELECOMMUNICATION INFORMATION:  
TELEPHONE: 716-263-1606  
TELEFAX: 716-263-1600  
INFORMATION FOR SEQ ID NO: 44:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 429 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-795-876-44

Query Match 11.5%; Score 60; DB 4; Length 429;  
Best Local Similarity 20.8%; Pred. No. 18;  
Matches 16; Conservative 15; Mismatches 26; Indels 20; Gaps 2;

OY 4 HVGKROHKEVTFW---ETDLSNNKTIVSLKKRPFLYCVIYIPVPLIILFDIAFI 60  
DB 232 HAITRAVYFNDNCWLSVET-----HLIYIHGPVMAALVNVFFLLNI 274  
OY 61 PKSLISOFQNNHYTHNH 77  
DB 275 VRVLYTKMRETHEASH 291

## RESULT 14

US-08-453-742-2  
Sequence 2, Application US/08453742  
Patent No. 5622839

GENERAL INFORMATION:  
APPLICANT: Moore, Emma E  
APPLICANT: Sheppard, Paul O  
APPLICANT: Kuestner, Rolf E  
TITLE OF INVENTION: Human Calcitonin Receptor  
NUMBER OF SEQUENCES: 27  
CORRESPONDENCE ADDRESS:

ADDRESSEE: TOWNSEND AND TOWNSEND KHOURIE and CREW  
STREET: One Market Plaza, Stewart St. Tower,  
STREET: Twentieth Floor  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94105-1492

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/453,742  
FILING DATE:

CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/100,887  
FILING DATE:

APPLICATION NUMBER: US 07/954,804  
FILING DATE: 30-SEP-1992  
ATTORNEY/AGENT INFORMATION:

NAME: Parmelee, Steven W  
REGISTRATION NUMBER: 31,990  
REFERENCE/DOCKET NUMBER: 13952-15-1

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-467-9600  
TELEFAX: 415-543-5043

INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 474 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-453-742-2

Query Match 11.5%; Score 60; DB 1; Length 474;

Best Local Similarity 20.8%; Pred. No. 20;  
Matches 16; Conservative 15; Mismatches 26; Indels 20; Gaps 2;  
OY 4 HVGKROHKEVTFW---ETDLSNNKTIVSLKKRPFLYCVIYIPVPLIILFDIAFI 60  
DB 277 HAITRAVYFNDNCWLSVET-----HLIYIHGPVMAALVNVFFLLNI 319  
OY 61 PKSLISOFQNNHYTHNH 77  
DB 320 VRVLYTKMRETHEASH 336

## RESULT 15

US-08-454-464-2  
Sequence 2, Application US/08454464  
Patent No. 5674689

GENERAL INFORMATION:

APPLICANT: Moore, Emma E  
APPLICANT: Sheppard, Paul O  
APPLICANT: Kuestner, Rolf E

TITLE OF INVENTION: Human Calcitonin Receptor  
NUMBER OF SEQUENCES: 27  
CORRESPONDENCE ADDRESS:

ADDRESSEE: TOWNSEND AND TOWNSEND KHOURIE and CREW  
STREET: One Market Plaza, Stewart St. Tower,  
STREET: Twentieth Floor  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94105-1492

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/454,464  
FILING DATE: 30-MAY-1995

CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/100,887  
FILING DATE: 02-AUG-1993

APPLICATION NUMBER: US 07/954,804  
FILING DATE: 30-SEP-1992  
ATTORNEY/AGENT INFORMATION:

NAME: Parmelee, Steven W  
REGISTRATION NUMBER: 31,990  
REFERENCE/DOCKET NUMBER: 13952-15-1  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 206-467-9600  
TELEFAX: 415-543-5043  
INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:  
LENGTH: 474 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-454-464-2

Query Match 11.5%; Score 60; DB 1; Length 474;  
Best Local Similarity 20.8%; Pred. No. 20;  
Matches 16; Conservative 15; Mismatches 26; Indels 20; Gaps 2;

OY 4 HVGKROHKEVTFW---ETDLSNNKTIVSLKKRPFLYCVIYIPVPLIILFDIAFI 60  
DB 277 HAITRAVYFNDNCWLSVET-----HLIYIHGPVMAALVNVFFLLNI 319  
OY 61 PKSLISOFQNNHYTHNH 77  
DB 320 VRVLYTKMRETHEASH 336

Search completed: April 14, 2003, 23:06:48

Tue Apr 15 15:15:18 2003

us-09-830-244b-1.rai

Job time : 91 secs

Tue Apr 15 15:15:19 2003

us-09-830-244b-1.rapb

Page 1

GenCore version 5.1.4-P5-4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2003, 23:03:36 ; Search time 163 Seconds

(without alignments)  
35.631 Million cell updates/sec

Title: US-09-830-244b-1

Perfect score: 522

Sequence: 1 MAFKVGKGRKHKFTFWETD.....NHTNHTNTRINRINISNCRP 95

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 248812 seqs, 61136040 residues

Total number of hits satisfying chosen parameters: 248812

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: /cn2\_6/ptodata/1/pubppa/US08\_NEW\_PUB pep.\*  
2: /cn2\_6/ptodata/1/pubppa/PCT\_NEW\_PUB pep.\*  
3: /cn2\_6/ptodata/1/pubppa/US06\_NEW\_PUB pep.\*  
4: /cn2\_6/ptodata/1/pubppa/US07\_PUBCOMB pep.\*  
5: /cn2\_6/ptodata/1/pubppa/US07\_PUBCOMB pep.\*  
6: /cn2\_6/ptodata/1/pubppa/US07\_PUBCOMB pep.\*  
7: /cn2\_6/ptodata/1/pubppa/US07\_PUBCOMB pep.\*  
8: /cn2\_6/ptodata/1/pubppa/US08\_PUBCOMB pep.\*  
9: /cn2\_6/ptodata/1/pubppa/US08\_PUBCOMB pep.\*  
10: /cn2\_6/ptodata/1/pubppa/US09\_PUBCOMB pep.\*  
11: /cn2\_6/ptodata/1/pubppa/US10\_PUBCOMB pep.\*  
12: /cn2\_6/ptodata/1/pubppa/US10\_PUBCOMB pep.\*  
13: /cn2\_6/ptodata/1/pubppa/US60\_NEW\_PUB pep.\*  
14: /cn2\_6/ptodata/1/pubppa/US60\_PUBCOMB pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	73.5	14.1	234	US-09-791-279-137	Sequence 137, App
2	73	14.0	412	US-09-841-132-362	Sequence 362, App
3	73	14.0	426	US-09-841-132-573	Sequence 573, App
4	73	14.0	433	US-09-841-132-573	Sequence 573, App
5	66	12.6	2150	US-10-135-322-17	Sequence 363, App
6	63.5	12.2	2787	US-09-893-238-15	Sequence 17, App
7	63	12.1	392	US-09-815-242-12755	Sequence 15, App
8	63	12.1	392	US-09-815-242-12755	Sequence 12755, A
9	63	12.1	414	US-09-815-242-13117	Sequence 13117, A
10	62	11.9	121	US-09-815-242-5780	Sequence 5780, App
11	62	11.9	367	US-09-815-242-118	Sequence 118, App
12	62	11.9	1092	US-09-815-242-10676	Sequence 10676, A
13	61	11.7	961	US-09-423-126-5	Sequence 5, App
14	61	11.7	1427	US-09-801-368-1332	Sequence 132, App
15	60.5	11.6	448	US-09-801-368-354	Sequence 354, App
16	60.5	11.6	467	US-09-361-443-4	Sequence 4, App
17	60	11.5	157	US-09-361-443-2	Sequence 2, App
18	60	11.5	490	US-09-881-752A-200	Sequence 200, App
19	60	11.5	666	US-09-801-368-6	Sequence 8, App

20	59.5	11.4	117	10	US-09-829-482-26	Sequence 26, App
21	58.5	11.2	235	9	US-09-866-050A-725	Sequence 725, App
22	58.5	11.2	1164	10	US-09-870-122-1	Sequence 1, App
23	58	11.1	99	9	US-10-183-119-2	Sequence 2, App
24	58	11.1	100	10	US-09-794-975-4	Sequence 4, App
25	58	11.1	103	10	US-09-892-475-2	Sequence 2, App
26	58	11.1	103	10	US-09-892-475-2	Sequence 2, App
27	58	11.1	117	9	US-09-422-569-10	Sequence 10, App
28	58	11.1	117	10	US-09-794-975-6	Sequence 6, App
29	58	11.1	117	10	US-09-823-153-2	Sequence 2, App
30	58	11.1	355	10	US-09-794-975-13	Sequence 13, App
31	58	11.1	695	10	US-09-794-975-10	Sequence 10, App
32	58	11.1	695	10	US-09-794-975-12	Sequence 12, App
33	58	11.1	695	10	US-09-794-975-10	Sequence 10, App
34	58	11.1	695	10	US-09-794-975-10	Sequence 10, App
35	58	11.1	695	10	US-09-794-975-10	Sequence 10, App
36	58	11.1	695	10	US-09-794-975-10	Sequence 10, App
37	58	11.1	695	10	US-09-794-975-10	Sequence 10, App
38	58	11.1	695	10	US-09-794-975-10	Sequence 10, App
39	58	11.1	695	10	US-09-794-975-10	Sequence 10, App
40	58	11.1	695	10	US-09-794-975-10	Sequence 10, App
41	58	11.1	695	10	US-09-794-975-10	Sequence 10, App
42	58	11.1	695	10	US-09-794-975-10	Sequence 10, App
43	58	11.1	695	10	US-09-794-975-10	Sequence 10, App
44	58	11.1	695	10	US-09-794-975-10	Sequence 10, App
45	58	11.1	695	10	US-09-794-975-10	Sequence 10, App

# ALIGNMENTS

RESULT 1  
US-09-791-279-137

Sequence 137, Application US/09791279

Publication No. US20030050456A1

GENERAL INFORMATION:

APPLICANT: Vogell, Gabriel

APPLICANT: Wood, Linda S.

APPLICANT: Lind, Peter

TITLE OF INVENTION: No. US20030050456A1 G Protein-Coupled Receptors

FILE REFERENCE: 00048, US1

CURRENT APPLICATION NUMBER: US/09/791,279

CURRENT FILING DATE: 2001-02-23

PRIOR APPLICATION NUMBER: 60/184,715

PRIOR FILING DATE: 2000-02-24

PRIOR APPLICATION NUMBER: 60/184,712

PRIOR FILING DATE: 2000-02-24

PRIOR APPLICATION NUMBER: 60/184,606

PRIOR FILING DATE: 2000-02-24

PRIOR APPLICATION NUMBER: 60/184,602

PRIOR FILING DATE: 2000-02-24

PRIOR APPLICATION NUMBER: 60/184,689

PRIOR FILING DATE: 2000-02-24

PRIOR APPLICATION NUMBER: 60/184,690

PRIOR FILING DATE: 2000-02-24

PRIOR APPLICATION NUMBER: 60/184,716

PRIOR FILING DATE: 2000-02-24

NUMBER OF SEQ ID NOS: 220

SOFTWARE: PatentIn version 3.0

SEQ ID NO 137

LENGTH: 234

TYPE: PRT

ORGANISM: Homo sapiens

US-09-791-279-137

Query Match 14.1% Score 73.5; DB 9; Length 234;  
Best Local Similarity 22.1% Pred. No. 0.45;  
Matches 21; Conservative 22; Mismatches 37; Indels 15; Gaps 3;

QY 12 EKVTEWETDLSNNKTLVSLKKRPPHL-----YCVIYI-PLVPLILFLFDIFPKSLISQ 67  
DB 133 FEMVLLKPELNSVSYLNKKEKRTITLIPSCVITFLCLIRSVFLFSPHKNCVSP 192  
QY 68 -----FQNNHYTHNHTNNTNIRNIIIS 91  
DB 193 LLSLNPILWPKHQRHAIHAHCEPQVOYCLIS 227

RESULT 2  
US-09-841-132-362  
Sequence 362, Application US/09841132  
Patent No. US20020061848A1  
GENERAL INFORMATION:  
APPLICANT: Bhalla, Ajay  
APPLICANT: Skelky, Yasir A.W.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND  
FILE REFERENCE: 210121.469C8  
CURRENT FILING DATE: 2001-04-23  
CURRENT APPLICATION NUMBER: US/09/841.132  
NUMBER OF SEQ ID NOS: 599  
SOFTWARE: FastSeq for Windows Version 3.0/4.0  
SEQ ID NO 362  
LENGTH: 412  
TYPE: PRT  
ORGANISM: Chlamydia  
US-09-841-132-362

Query Match 14.0% Score 73; DB 10; Length 412;  
Best Local Similarity 26.5% Pred. No. 1;  
Matches 30; Conservative 19; Mismatches 34; Indels 30; Gaps 7;

QY 2 WFH--KVGR-----KQFKVTEWETDLSNNKTLVSLKKRPPHL-----YCVIYIPLV 47  
DB 95 WEHTKQAGRYSLFEKLSFRAS-----SSSEILIEKPEPQLAILASPFVAVYRPN 148  
QY 48 PKLILFLDIA-FIPKSLIS-----QFQNNHYTHNHTNNTNIRNIIISNCR 95  
DB 149 P-----FLSSGPFMPKTYVQGTLLVQKNPYYDDAHAVELHSIDFRIIPNIYT 196

RESULT 3  
US-09-841-132-573  
Sequence 573, Application US/09841132  
Patent No. US20020061848A1  
GENERAL INFORMATION:  
APPLICANT: Bhalla, Ajay  
APPLICANT: Skelky, Yasir A.W.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND  
FILE REFERENCE: 210121.469C8  
CURRENT FILING DATE: 2001-04-23  
CURRENT APPLICATION NUMBER: US/09/841.132  
NUMBER OF SEQ ID NOS: 599  
SOFTWARE: FastSeq for Windows Version 3.0/4.0  
SEQ ID NO 573  
LENGTH: 426  
TYPE: PRT  
ORGANISM: C. Trachomatis D serovar  
US-09-841-132-573

Query Match 14.0% Score 73; DB 10; Length 426;  
Best Local Similarity 26.5% Pred. No. 1;  
Matches 30; Conservative 19; Mismatches 34; Indels 30; Gaps 7;

QY 2 WFH--KVGR-----KQFKVTEWETDLSNNKTLVSLKKRPPHL-----YCVIYIPLV 47  
DB 109 WEHTKQAGRYSLFEKLSFRAS-----SSSEILIEKPEPQLAILASPFVAVYRPN 162  
QY 48 PKLILFLDIA-FIPKSLIS-----QFQNNHYTHNHTNNTNIRNIIISNCR 95  
DB 163 P-----FLSSGPFMPKTYVQGTLLVQKNPYYDDAHAVELHSIDFRIIPNIYT 210

RESULT 4  
US-09-841-132-363  
Sequence 363, Application US/09841132  
Patent No. US20020061848A1  
GENERAL INFORMATION:  
APPLICANT: Bhalla, Ajay  
APPLICANT: Skelky, Yasir A.W.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND  
FILE REFERENCE: 210121.469C8  
CURRENT FILING DATE: 2001-04-23  
CURRENT APPLICATION NUMBER: US/09/841.132  
NUMBER OF SEQ ID NOS: 599  
SOFTWARE: FastSeq for Windows Version 3.0/4.0  
SEQ ID NO 363  
LENGTH: 433  
TYPE: PRT  
ORGANISM: Chlamydia  
US-09-841-132-363

Query Match 14.0% Score 73; DB 10; Length 433;  
Best Local Similarity 26.5% Pred. No. 1.1;  
Matches 30; Conservative 19; Mismatches 34; Indels 30; Gaps 7;

QY 2 WFH--KVGR-----KQFKVTEWETDLSNNKTLVSLKKRPPHL-----YCVIYIPLV 47  
DB 116 WEHTKQAGRYSLFEKLSFRAS-----SSSEILIEKPEPQLAILASPFVAVYRPN 169  
QY 48 PKLILFLDIA-FIPKSLIS-----QFQNNHYTHNHTNNTNIRNIIISNCR 95  
DB 170 P-----FLSSGPFMPKTYVQGTLLVQKNPYYDDAHAVELHSIDFRIIPNIYT 217

RESULT 5  
US-10-135-322-17  
Sequence 17, Application US/10135322  
Patent No. US20020173017A1  
GENERAL INFORMATION:  
APPLICANT: BENEFY, PN  
APPLICANT: HELARIUTTA, Y  
APPLICANT: MAHONEN, AP  
APPLICANT: BONKE, AMM  
APPLICANT: KAUPPINEN, L  
APPLICANT: KIIKONEN, M  
TITLE OF INVENTION: WOODEN LEG GENE, PROMOTER AND USES THEREOF  
FILE REFERENCE: 5914-0066-999  
CURRENT APPLICATION NUMBER: US/10/135.322  
CURRENT FILING DATE: 2002-04-30  
PRIOR APPLICATION NUMBER: 60/253,739  
NUMBER OF SEQ ID NOS: 43  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 17  
LENGTH: 2150  
TYPE: PRT  
ORGANISM: Arabidopsis thaliana  
US-10-135-322-17

Query Match 12.6% Score 66; DB 9; Length 2150;  
Best Local Similarity 44.8% Pred. No. 46;  
Matches 13; Conservative 5; Mismatches 11; Indels 0; Gaps 0;  
QY 66 SQFQNNHYTHNHTNNTNIRNIIISNCR 94





```

Query Match      11.9%  Score 62:  DB 12;  length 121.
Best Local Similarity 23.9%:  Pred. No. 4,7;
Matches 21;  Conservative 14;  Mismatches 29;  Indels 24;  Gaps 4

QY      17  WEDTDSNKKTVLSVKKR---KPEHL-YGVYIPLVP-----KLLFLDLDAFIRKSL 64
      11  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      37  WEQTGTGNDLNPALYSKSYDLRFLFDQCEHLYKALFPEEDYKDSPELDLHVLGDLT--- 92

QY      65  ISQFQNNHYTHNTNTNNTNNIRFNITSN 92

Db      93  -----GSHQDQNNRTEDIALSCLN 112

```

```

QY 17 WEDTSSNKKVLTAKK---KRFH-YGVYIPLVP-----KLIIFLDIAFIRKSL 64
   11 : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 37 WEQGTGNDIMPKLSTYLPFFHQAQCFICALFEDYKFSDELHIMIGDIL---- 92
   31 : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 65 ISQFQNNHYTHHTNHNINIRFIITSN 92
   61 : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 93 -----QSHODQNRKEDIALSLNH 112
   91 : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 11
US-09-815-242-10676
; Sequence 10676, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10676
; LENGTH: 367
; TYPE: PRF
; ORGANISM: Enterococcus faecalis
US-09-815-242-10676

```

RESULT 12  
US-09-423-126-5

```

1  APPLICANT: Buchter-Larsen, et al.
2  TITLE OF INVENTION: A PROCESS OF PREPARING AN ANTI-OXIDANT
3  FILE REFERENCE: 674509-2020
4  CURRENT APPLICATION NUMBER: US/09/423,126
5  CURRENT FILING DATE: 1999-11-05
6  PRIOR APPLICATION NUMBER: PCT/IB98/00708
7  PRIOR FILING DATE: 1998-05-06
8  PRIOR APPLICATION NUMBER: GB 9709161.5
9  PRIOR FILING DATE: 1997-05-06
10 NUMBER OF SEQ ID NOS: 12
11 SOFTWARE: PatentIn version 3.0
12
13 SEQ ID NO 5
14
15 LENGTH: 1092
16
17 TYPE: prt
18
19 ORGANISM: Unknown
20
21 FEATURE:
22
23 OTHER INFORMATION: fungus sp. or fungus infected gracilariopsis sp
24
25 US-09-423-126-3

```

```

Query Match      11.9%; Score 62; DB 9; Length 1092;
Best Local Similarity 23.1%; Pred. No. 61;
Matches 27; Conservative 15; Mismatches 29; Indels 46; Gaps 6;

QY      6  VGRKQHEKTVEMFTDSSNNKTVLSKKRP--FHLYCYIYPLPKLILFLDIATPKS 63
      | | : : : | | : : | : : | : : | : : | : : | : : | : : |
Db      40  VRSKKHNVSM--TALSDKQTALISGPDNPDCINQNDYIPVA-----GTPPLS 87
      | | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY      64  -----LISOFNNHYTHNH-----TNNTNTN--IREN 88
      | | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db      88  NFNWYAAGSSTPGGITDWTATGNVAFEDLIDNPSTSNNHFPQVIOYVTSNNNSFEIREN 144

```

RESULT 13  
 US-09-801-368-132  
 ? Sequence 132, Application US/09801368  
 ? Patent No. US20020128250A1  
 ? GENERAL INFORMATION:  
 ? APPLICANT: Busby, Robert  
 ? APPLICANT: Call, Brian  
 ? APPLICANT: Hecht, Peter  
 ? APPLICANT: Holtzman, Doug  
 ? APPLICANT: Madden, Kevin  
 ? APPLICANT: Maxon, Mary  
 ? APPLICANT: Milne, Todd  
 ? APPLICANT: No. US20020128250A1man, Thea  
 ? APPLICANT: Royer, John  
 ? APPLICANT: Salama, Sofie  
 ? APPLICANT: Sherman, Amir  
 ? APPLICANT: Silva, Jeff  
 ? APPLICANT: Summers, Eric  
 ? TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi  
 ? FILE REFERENCE: 109272.147  
 ? CURRENT APPLICATION NUMBER: US/09/801,368  
 ? PRIOR FILING DATE: 2001-03-07  
 ? PRIOR APPLICATION NUMBER: US 09/487,558  
 ? PRIOR FILING DATE: 2000-01-19  
 ? PRIOR APPLICATION NUMBER: US 60/160,587  
 ? PRIOR FILING DATE: 1999-10-20  
 ? NUMBER OF SEQ ID NOS: 440

```

; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 132
; LENGTH: 961
; TYPE: PR1
; ORGANISM: Saccharomyces cerevisiae
us-09-801-368-132

```

Query Match	11.7%	Score 61;	DB 10;	Length 961;
Best Local Similarity	43.5%;	Pred. No. 69;		
Matches 10;	Conservative	8;	Mismatches	5; Indels 0; Gaps 0;

```
QY 70 NNHYTHNHTNHTNNIRENIISN 92
    ||: :|: |:|:|: |: :|
Db 533 NNNNNNNNNNNNNNNIKKNVDNN 555
```

```

: RESULT 14
: US-09-801-368-354
: Sequence 354, Application US/09801368
: Patent No. US20020128250A1
: GENERAL INFORMATION:
: APPLICANT: Busby, Robert
: APPLICANT: Cali, Brian
: APPLICANT: Hecht, Peter
: APPLICANT: Holtzman, Doug
: APPLICANT: Madden, Kevin
: APPLICANT: Maxon, Mary
: APPLICANT: Milne, Todd
: APPLICANT: No. us20020128250A1man, Thea
: APPLICANT: Royer, John
: APPLICANT: Salama, Sofie
: APPLICANT: Sherman, Amir
: APPLICANT: Silva, Jeff
: APPLICANT: Summers, Eric
: TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
: FILE REFERENCE: 109272.147
: CURRENT APPLICATION NUMBER: US/09/801.368
: CURRENT FILING DATE: 2001-03-07
: PRIOR APPLICATION NUMBER: US 09/487.558
: PRIOR FILING DATE: 2000-01-19
: PRIOR APPLICATION NUMBER: US 60/160.587
: PRIOR FILING DATE: 1999-10-20
: NUMBER OF SEQ ID NOS: 440
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 354
: LENGTH: 1427
: TYPE: PRT
: ORGANISM: Saccharomyces cerevisiae
: US-09-801-368-354

```

```

Query Match      11.7%; Score 61; DB 10; Length 1427;
Best Local Similarity 26.7%; Pred. No. 1,1e+02;
Matches 23; Conservative 13; Mismatches 34; Indels 16; Gaps

QY      11 HKVETW-----ETDLSNKKTLVSLKKRPFLHYCVIYIPVLPKLIIFLDIAFPKSL 64
      | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      283 HILIMFNIDICQIDTNAPVAATITS-SQKEPF-----LVTKITDMLLHKRYIVSSS 333
      | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      65 ISOFOHHYTHNTNTNTNNIRENT 90
      | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      334 KSMINDENTYINDIKKN-NKIKLNTIL 358

RESULT 15
US-09-361-443-4
; Sequence 4, Application US/09361443
; Patent No. US20020150591A1
; GENERAL INFORMATION:
; APPLICANT: Murdin, Andrew D
; APPLICANT: Cohen, Raymond P
; TITLE OF INVENTION: CHLAMYDIA ANTIGENS AND CORRESPONDING DNA FRAGMENTS AND
; FILE REFERENCE: 19721-005 (AV-5)

```

us-09-830-244b-1.rapb

Page 6

Query Match	11.6%	Score 60.5	DB 10,	Length 448,
Best Local Similarity	29.3%	Pred. No. 32,		
Matches	24;	Conservative	14;	Mismatches 23;
				Indels 21;
				Gaps 5;

[illegible]

Search completed: April 14, 2003, 23:09:45  
Job time : 164 secs



AL512306/c  
 LOCUS AL512306 133984 bp DNA linear PRI 01-FEB-2002  
 DEFINITION Human DNA sequence from clone RP11-430C7 on chromosome 1, complete  
 sequence.  
 ACCESSION AL512306  
 VERSION AL512306.16 GI:18491332  
 KEYWORDS HTG.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 REFERENCE 1 (bases 1 to 133984)  
 AUTHORS Bray-Allen, S.  
 TITLE Direct Submission  
 JOURNAL Submitted (01-FEB-2002) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 humquer@sanger.ac.uk  
 COMMENT On Feb 4, 2002 this sequence version replaced gi:18477311.  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.  
 This sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone or more than one M13 subclone; and the  
 assembly was confirmed by restriction digest. The following  
 abbreviations are used to associate primary accession numbers given  
 in the feature table with their source databases: Em: EMBL; SW: SWISSPROT; Tr: TrEMBL; Wp: WormPeP; Information on the WormPeP  
 database can be found at  
 http://www.sanger.ac.uk/Projects/C\_elegans/wormpep  
 This sequence was generated from part of bacterial clone contigs of human  
 chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping  
 group. Further information can be found at  
 http://www.sanger.ac.uk/HGP/Ch1  
 RP11-430C7 is from the library RP11-11.2 constructed by the group  
 of Pletier de Jong. For further details see  
 http://www.chori.org/bacpac/home.htm  
 VECTOR: pBACe3.6  
 IMPORTANT: This sequence is not the entire insert of clone  
 RP11-430C7. It may be shorter because we sequence overlapping  
 sections only once, except for a short overlap.  
 The true right end of clone RP11-430C7 is at 133984 in this  
 sequence. The true left end of clone RP11-739N20 is at 97715 in  
 this sequence. The true right end of clone RP11-2317 is at 2000 in  
 this sequence.

FEATURES  
 source  
 1..133984  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="1"  
 /clone="RP11-430C7"  
 /clone\_lib="RP11-11.2"  
 123904..123933  
 /note="Single clone region. Reads generated from a  
 transposon library derived from a single pUC clone.  
 Restriction digest data confirm the assembly."

BASE COUNT 37368 a 32097 c 30202 g 34317 t

Alignment Scores:  
 Pired. No.: 8,83e-28 Length: 133984  
 Score: 349.00 Matches: 65  
 Percent Similarity: 100.00% Conservative: 1  
 Best Local Similarity: 98.48% Mismatches: 0  
 Query Match: 66.86% Indels: 0  
 DB: 9 Gaps: 0

US-09-830-244b-1 (1-95) x AL512306 (1-133984)  
 QY 1 MettPphenHsLysValGlyArgLysGlnHisPheLysValThrPheTrpGluTrpAsp 20  
 Db 91100 ATGTGGTTTCATTAAGGGGAGAAAAACAGCATTTTAAAGTAACCTTTTGGGAGACTGAT 91041  
 QY 21 LeuSerAsnAsnLysThrLeuValSerLeuLysLysLysLysProPheHisLeuTrpCys 40  
 Db 91040 TTGAGTAATAATAAAGCTGCTGCTCCCTTAAGAAAAAAACCTTCACCTTACTGT 90981  
 QY 41 ValIleTyrIleProLeuValProLysLeuIleIleLeuPheLeuAspIleAlaPheIle 60  
 Db 90980 GTCATTATATACCCCTTACTGCACAAAGTAATATCTTATTCGTGATATTCCTTTATA 90921  
 QY 61 PolySerLeuIleSer 66  
 Db 90920 CCAAAGACCTTATTCAGC 90903

RESULT 2  
 AC021462 153023 bp DNA linear HTG 03-APR-2000  
 LOCUS Homo sapiens clone RP11-276C1, WORKING DRAFT SEQUENCE, 21 unordered  
 DEFINITION pieces.  
 ACCESSION AC021462.3 GI:7387343  
 VERSION HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 KEYWORDS Homo sapiens.  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 REFERENCE 1 (bases 1 to 153023)  
 AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
 TITLE Homo sapiens, clone RP11-276C1  
 JOURNAL Unpublished  
 AUTHORS 2 (bases 1 to 153023)  
 REFERENCES 2 (bases 1 to 153023)  
 JOURNAL Unpublished

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
 Anderson, S., Baldwin, J., Barna, N., Beckert, R., Beda, F.,  
 Boguski, M., Boukhalil, B., Brown, A., Burkett, G., Castle, A.,  
 Chovel, X., Colangelo, M., Collins, S., Collins, A., Cooke, P.,  
 DeArnell, K., Dewar, K., Dominko, M., Doyle, M., Fenebor, J.,  
 Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J.,  
 Gardina, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,  
 Howland, J., Johnson, R., Jones, C., Kann, L., Karas, A., Klein, J.,  
 Landers, T., Lehotzky, J., Levine, R., Liu, C., Liu, G., Locke, K.,  
 MacDonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K.,  
 McNeeters, R., Melidini, J., Meneus, L., Morrow, J., Naylor, J.,  
 Norman, C., O'Connor, T., O'Donnell, P., Oliver, T., Peterson, K.,  
 Pletier, N., Pisan, C., Pollard, V., Raymond, C., Riley, R., Rothman, D.,  
 Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,  
 Stojanovic, N., Subramanian, A., Talamas, J., Testa, S., Theodore, J.,  
 Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W., J.,  
 Zimmer, A., and Zody, M.  
 Direct Submission  
 Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Apr 3, 2000 this sequence version replaced gi:7230200.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

TITLE  
 JOURNAL  
 COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: MIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 Project Information  
 Center project name: L5201  
 Center clone name: 276.C.1  
 Summary Statistics  
 Sequencing vector: M13, M77815, 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 141815 bases at least Q40  
 Consensus quality: 147368 bases at least Q30

Consensus quality: 149481 bases at least Q20  
 Insert size: 15700; agarose-fp  
 Insert size: 151023; sum-of-contigs  
 Quality coverage: 4.3 in Q20 bases; agarose-fp  
 Quality coverage: 4.5 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 21 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1 1015: contig of 1015 bp in length
1016 1115: gap of 100 bp
1116 2276: contig of 1161 bp in length
2277 2376: gap of 100 bp
2377 4009: contig of 1633 bp in length
4010 4109: gap of 100 bp
4110 5737: contig of 1628 bp in length
5738 5837: gap of 100 bp
5838 6939: contig of 1102 bp in length
6940 7039: gap of 100 bp
7040 9504: contig of 2465 bp in length
9505 9604: gap of 100 bp
11735 11834: contig of 2130 bp in length
11835 14111: contig of 2277 bp in length
14112 14211: gap of 100 bp
14212 16973: contig of 2762 bp in length
16974 17073: gap of 100 bp
17074 19588: contig of 2515 bp in length
19589 19688: gap of 100 bp
19689 22275: contig of 2587 bp in length
22276 22375: gap of 100 bp
22376 25823: contig of 3448 bp in length
25824 25923: gap of 100 bp
25924 31307: contig of 5384 bp in length
31308 31407: gap of 100 bp
31408 34863: contig of 3456 bp in length
34864 34963: gap of 100 bp
34964 41382: contig of 6419 bp in length
41383 41482: gap of 100 bp
41483 49024: contig of 7542 bp in length
49025 49124: gap of 100 bp
49125 58672: contig of 9548 bp in length
58673 58772: gap of 100 bp
58773 69622: contig of 10850 bp in length
69623 69722: gap of 100 bp
69723 88191: contig of 18469 bp in length
88192 88291: gap of 100 bp
88292 107084: contig of 18793 bp in length
107085 107184: gap of 100 bp
107185 153023: contig of 45839 bp in length.
Location/Qualifiers
1. 153023
/Organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-276C1"
/clone_1lb="RPCT-11 Human Male BAC"
1. 1015
misc_feature
/note="assembly-fragment"
1116. 2276
/note="assembly-fragment"
2377. 4009
/note="assembly-fragment"
4110. 5737
/note="assembly-fragment"
5838. 6939
/note="assembly-fragment"
7040. 9504
/note="assembly-fragment"

```

```

misc_feature 9605..11734
/note="assembly-fragment"
misc_feature 11835..14111
/note="assembly-fragment"
misc_feature 14212..16973
/note="assembly-fragment"
misc_feature 17074..19588
/note="assembly-fragment"
misc_feature 19689..22275
/note="assembly-fragment"
misc_feature 22376..25823
/note="assembly-fragment"
misc_feature 25924..31307
/note="assembly-fragment"
misc_feature 31408..34863
/note="assembly-fragment"
misc_feature 34964..41382
/note="assembly-fragment"
misc_feature 41483..49024
/note="assembly-fragment"
misc_feature 49125..58672
/note="assembly-fragment"
misc_feature 58773..69622
/note="assembly-fragment"
misc_feature 69723..88191
/note="assembly-fragment"
misc_feature 88292..107084
/note="assembly-fragment"
misc_feature 107185..153023
/note="assembly-fragment"

```

BASE COUNT 41494 a 35192 c 34573 g 39756 t 2008 others  
 ORIGIN  
 vector:side:left"

Alignment Scores:  
 Pred. No.: 1.01e-27 Length: 153023  
 Score: 349.00 Matches: 65  
 Percent Similarity: 100.00% Consensitive: 1  
 Best Local Similarity: 98.488 Mismatches: 0  
 Query Match: 66.868 Indels: 0  
 DB: 2 Gaps: 0

US-09-830-244b-1 (1-95) x AC021462 (1-153023)

```

QY 1 MetTrrPheHtIsYsValGlyArgLysGlnHisPheLysValAlhrrPheTrgLuThrAsp 20
|||||
DB 442 ATGTGTTTCATTAAGTGGAGAGAAACAGCATTTTAAAGTAACTTTTGGAGACTGAT 501
|||||
QY 21 LeuSerAsnAsnLysThrLeuValSerLeuLysLysLysLysProPheHisLeuTyrcys 40
|||||
DB 502 TTGAGTAATAATTAACCTGCGTCCCTTAAGAAAAAACCTTCCACCTTTACTGT 561
|||||
QY 41 ValIleTyrlleProLeuValProLysLeuIleIleLeuPheLeuAspIleAlaPheIle 60
|||||
DB 562 GTCATTATATCCCTTACCTTCCAAAGTAAATATCTATTCTCGATATGCTTTTAA 621
|||||
QY 61 ProLysSerLeuIleSer 66
|||||
DB 622 CCAAGACCCCTTATCAGC 639
|||||

```

RESULT 3  
 AL450424 146805 bp DNA linear HTG 12-SEP-2001  
 LOCUS Homo sapiens chromosome 1 clone RP11-561I16.\*\*\* SEQUENCING IN  
 DEFINITION PROGRESS \*\*\* 25 unordered pieces.  
 ACCESSION AL450424.3 GI:12331147  
 VERSION HTG: HTGS\_PHASE1; HTGS\_CANCELLED.  
 KEYWORDS human.  
 SOURCE  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.



REFERENCE  
AUTHORS  
TITLE  
JOURNAL

## COMMENT

1 (bases 1 to 146805)  
McLay, K.  
Submitted (11-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
On Jan 22, 2001 this sequence version replaced g1:11493361.  
-----  
Center: Sanger Centre  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: humquerry@sanger.ac.uk  
-----  
Project Information  
Center project name: BA563116  
-----  
Summary Statistics  
Sequencing program: XGAP4; version 4.5  
Sequencing vector: plasmid; L08752; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Consensus quality: 135323 bases at least Q40  
Consensus quality: 139636 bases at least Q30  
Consensus quality: 141952 bases at least Q20  
Insert size: 14405; sum-of-contigs  
Insert size: 171032; 2.28 error; agarose-fp  
Quality coverage: 3.60x in Q20 bases; sum-of-contigs Quality  
Coverage: 3.25x in Q20 bases; agarose-fp  
-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 25 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 5105: contig of 5105 bp in length  
\* 5106 5205: gap of 100 bp  
\* 5206 10570: contig of 5365 bp in length  
\* 10571 10670: gap of 100 bp  
\* 10671 17626: contig of 6956 bp in length  
\* 17627 17726: gap of 100 bp  
\* 17727 20478: contig of 2752 bp in length  
\* 20479 20578: gap of 100 bp  
\* 20579 23035: contig of 2457 bp in length  
\* 23036 23135: gap of 100 bp  
\* 23136 25308: contig of 2173 bp in length  
\* 25309 25408: gap of 100 bp  
\* 25409 30469: contig of 5061 bp in length  
\* 30470 30569: gap of 100 bp  
\* 30570 32688: contig of 2119 bp in length  
\* 32689 32788: gap of 100 bp  
\* 32789 36187: contig of 3399 bp in length  
\* 36188 36287: gap of 100 bp  
\* 36288 48988: contig of 12701 bp in length  
\* 48989 49088: gap of 100 bp  
\* 49089 67333: contig of 18245 bp in length  
\* 67334 67433: gap of 100 bp  
\* 67434 82317: contig of 14884 bp in length  
\* 82318 82417: gap of 100 bp  
\* 82418 84614: contig of 2197 bp in length  
\* 84615 84714: gap of 100 bp  
\* 84715 90326: contig of 5612 bp in length  
\* 90327 90426: gap of 100 bp  
\* 90427 92647: contig of 2221 bp in length  
\* 92648 92747: gap of 100 bp  
\* 92748 97690: contig of 4943 bp in length  
\* 97691 97790: gap of 100 bp  
\* 97791 100799: contig of 3009 bp in length  
\* 100800 100899: gap of 100 bp  
\* 100900 104202: contig of 3303 bp in length  
\* 104203 104302: gap of 100 bp  
\* 104303 107157: contig of 2855 bp in length  
\* 107158 107257: gap of 100 bp  
\* 107258 110300: contig of 3043 bp in length

FEATURES  
source

\* 110301 110400: gap of 100 bp  
\* 110401 113955: contig of 3555 bp in length  
\* 113956 114055: gap of 100 bp  
\* 114056 118801: contig of 4746 bp in length  
\* 118802 118901: gap of 100 bp  
\* 118902 129831: contig of 10930 bp in length  
\* 129832 129931: gap of 100 bp  
\* 129932 137545: contig of 7614 bp in length  
\* 137546 137645: gap of 100 bp  
\* 137646 146805: contig of 9160 bp in length.  
Location/Qualifiers  
1. 146805  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/clone="RP11-563116"  
/clone\_id="RPC1-11.2"  
1. 5105  
/note="assembly-fragment:00368  
clone\_end:T7  
vector\_side:left"  
5206..10570  
/note="assembly-fragment:00594  
fragment\_chain:1"  
10671..17626  
/note="assembly-fragment:00786  
fragment\_chain:1"  
17727..20478  
/note="assembly-fragment:00844  
fragment\_chain:1"  
20579..23035  
/note="assembly-fragment:00011  
fragment\_chain:1"  
23136..25308  
/note="assembly-fragment:01029  
fragment\_chain:1"  
25409..30469  
/note="assembly-fragment:00148  
fragment\_chain:2"  
30570..32688  
/note="assembly-fragment:00961  
fragment\_chain:2"  
32789..36187  
/note="assembly-fragment:00310  
fragment\_chain:3"  
36288..48988  
/note="assembly-fragment:01150  
fragment\_chain:3"  
49089..67333  
/note="assembly-fragment:00444  
fragment\_chain:4"  
67434..82317  
/note="assembly-fragment:00140  
fragment\_chain:4"  
82418..84614  
/note="assembly-fragment:00657  
fragment\_chain:5"  
84715..90326  
/note="assembly-fragment:00757  
fragment\_chain:5"  
90427..92647  
/note="assembly-fragment:00677  
fragment\_chain:6"  
92748..97690  
/note="assembly-fragment:00248  
fragment\_chain:6"  
97791..100799  
/note="assembly-fragment:00091"  
100900..104202  
/note="assembly-fragment:00538"  
104303..107157  
/note="assembly-fragment:00675"  
107258..110300

```

misc_feature /note="assembly_fragment:00695"
110401..113955
misc_feature /note="assembly_fragment:01316"
114056..118801
misc_feature /note="assembly_fragment:01292"
118902..129831
misc_feature /note="assembly_fragment:00551"
129832..137545
misc_feature /note="assembly_fragment:01410"
137646..146805
misc_feature /note="assembly_fragment:00763"
fragment_chain:7
clone_end:SP6
vector_side:right"
BASE COUNT 37508 a 33240 c 34076 g 39557 t 2424 others
ORIGIN

```

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:
1	321.00	146805	64
Percent Similarity:	98.48%	Conservative:	1
Best Local Similarity:	96.97%	Mismatches:	1
Query Match:	61.49%	Indels:	1
DB:	2	Gaps:	0

```

US-09-830-244b-1 (1-95) x AL450424 (1-146805)
QY 1 MetTrpHisLysValAlaArgLysGlnHisPheLysValThrPheTrpGluThrAsp 20
DB 84902 ATGTGGTTTCATTAAGTGGAGAAACAGCATTTAAAGTAAAC-TTTGGGAGACTGAT 84960
QY 21 LeuSerAsnAsnLysThrLeuValSerLeuLysLysLysProPheHisLeuTyrCys 40
DB 84961 TTGAGTAATTAATTAACCTGTGCTCCCTTAAGAAAAAACCCCTTCACCTTACTGT 85020
QY 41 ValLeuTyrIleProLeuValProLysLeuIleIleLeuPheLeuAspIleAlaPheIle 60
DB 85021 GTCATTATATCCCTTACTGTTCAAGTTAATTAATTCATTCTGGAATATCTTTTATA 85080
QY 61 ProLysSerLeuIleSer 66
DB 85081 CCAAAGACCTTATCAGC 85098

RESULT 4
LOCUS HUMSTRNA 542 bp mRNA linear PRI 13-JAN-1995
DEFINITION Human statherin mRNA, complete cds.
ACCESSION M18371
VERSION M18371.1 GI:338610
KEYWORDS statherin.
SOURCE Human female submandibular gland, cDNA to mRNA, clone pBRHSMSP9B8.2.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 542)
AUTHORS Dickinson,D.P., Ridall,A.L. and Levine,M.J.
TITLE Human submandibular gland statherin and basic histidine-rich
peptide are encoded by highly abundant mRNA's derived from a common
ancestral sequence
JOURNAL Biochem. Biophys. Res. Commun. 149 (2), 784-790 (1987)
MEDLINE 86106506
PUBMED 3426601
COMMENT Draft entry and computer readable copy of sequence [1] kindly
provided by D.P. Dickinson 21-MAR-1988.
FEATURES
Source
1..342
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="4q11-q13"
1..542
gene

```

```

CDS
/gene="STATH"
57..245
/gene="STATH"
/note="statherin precursor"
/codon_start=1
/protein_id="AAA60600.1"
/db_xref="GI:338611"
/db_xref="GDB:G00-120-391"
/translation="MKELVFAFLLALNMSMGADSEKEFLRRIGFGYGPYPQVPV
EGPLYPQYQPYQYTF"
57..113
/gene="STATH"
/note="statherin signal peptide"
114..242
/gene="STATH"
/product="statherin"
mat_peptide
BASE COUNT 170 a 117 c 76 g 179 t
ORIGIN 262 bp upstream of PstI site; chromosome 4q11-q13.

```

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:
3	335.11	542	37
Percent Similarity:	178.50	Conservative:	7
Best Local Similarity:	63.77%	Mismatches:	8
Query Match:	53.62%	Indels:	17
DB:	9	Gaps:	2

```

US-09-830-244b-1 (1-95) x HUMSTRNA (1-542)
QY 27 LeuValSerLeuLysLysLysProPheHisLeuTyrCysValIleTyrIleProLeu 46
DB 112 CTAATTCATCTGAGAGAAAT-----TTTGGCTA----- 141
QY 47 ValProLysLeuIleIleLeuPheLeuAspIleAlaPheIleProLysSerLeuIleSer 66
DB 142 -----CAATTTGGAAGATTGCTTATGCTTATGCTTATCAGC 180
QY 67 GlnPheGlnAsnAsnHisTyrThrHisAsnHisThrAsnHisAsnThrAsnAsnIleArg 86
DB 181 CAGTCCAGAACCAACCATATACCAACCAACCAACCAACCAACCAACCAACCAATTAACCT 240
QY 87 PheAsnIleIleSerAsnGlyArgThr 95
DB 241 TTAATATCATCATCACTGACGACCA 267

RESULT 5
LOCUS HUMSTRNA 552 bp mRNA linear PRI 13-JAN-1995
DEFINITION Human statherin mRNA, complete cds.
ACCESSION M18078
VERSION M18078.1 GI:338507
KEYWORDS regulatory protein; statherin.
SOURCE Human parotid gland, cDNA to mRNA, clone H772B.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 552)
AUTHORS Sabatini,L.M., Carlock,L.R., Johnson,G.W. and Azen,E.A.
TITLE cDNA cloning and chromosomal localization (4q11-13) of a gene for
statherin, a regulator of calcium localization in saliva
JOURNAL Am. J. Hum. Genet. 41 (6), 1048-1060 (1987)
MEDLINE 86074310
PUBMED 3502720
COMMENT Draft entry and computer-readable sequence [1] kindly submitted by
L. Sabatini 19-JAN-1988.
FEATURES
Source
1..552
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="4q11-q13"
1..552
/gene="STATH"
73..261
CDS

```

```

/gene="STAT1"
/Note="statherin protein"
/codon_start=1
/protein_id="AAA60594.1"
/db_xref="GI:338508"
/db_xref="GDB:G00-120-391"
/translation="MKFLVFATLAWSMIGADSEERLRRIGRFGYGYPPVP
NPPLPPPOPOPOOYTP"
73.129
sig_peptide
/gene="STAT1"
/Note="statherin signal peptide"
BASE COUNT 172 a 122 c 78 g 180 t
ORIGIN 278 bp upstream of pstr site; chromosome 4q11-q13.

Alignment Scores:
Pred. No.: 3.42e-11 Length: 552
Score: 178.50 Matches: 37
Percent Similarity: 63.77% Conservative: 7
Best Local Similarity: 53.62% Mismatches: 8
Query Match: 34.20% Indels: 17
DB: 9 Gaps: 2

US-09-830-244b-1 (1-95) x HUMSTAT1A (1-552)
OY 27 LeuValSerLeuLysLysLysProPheHisLeuTyrCysValIleTyrIleProLeu 46
Db 128 CTGATTCACTCGAAGAGAAAT-----TTTCCGTA-----157
OY 47 ValProLysLeuIleIleLeuPheLeuAspIleAlaPheIleProLysSerLeuIleSer 66
Db 158 -----/-----GAATTGGAGATTCCGTTATGGGTATGGCCCTTATCAGC 196
OY 67 GluPheGlnAsnAsnHisTyrThrHisAsnHisThrAsnHisAsnThrAsnAsnIleArg 86
Db 197 CAGTTCCAGAACACCACTATACCCACACCATACCAACCACTACCAATATATACCT 256
OY 87 PheAsnIleIleSerAsnCysArgThr 95
Db 257 TTTAATATCATCACTACTGACGAGACA 283

RESULT 6
AK092678 1584 bp mRNA linear 11995
LOCUS AK092678 Homo sapiens cDNA FL35359 f1s, clone SALGL1000107, highly similar
DEFINITION AK092678 Homo sapiens cDNA FL35359 f1s, clone SALGL1000107, highly similar
ACCESSION AK092678
VERSION AK092678.1 GI:21751327
KEYWORDS oligo capping; f1s (full insert sequence)
SOURCE Homo sapiens salivary gland cDNA to mRNA, clone_11b:SALGL1
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1
REFERENCE
AUTHORS Oshima,A., Takahashi-Fujii,A., Tanase,T., Imose,N., Takeuchi,K.,
Arita,M., Musashino,K., Yuki,H., Hara,H., Sugiyama,T., Irie,R.,
Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y.,
Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H.,
Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K.,
Wagatsuma,M., Murakawa,K., Kanehori,K., Sugiyama,A., Kawakami,B.,
Suzuki,Y., Sugano,S., Nagahori,K., Masuno,Y., Nagai,K. and
Isogai,T.
NEO human cDNA sequencing project
unpublished
2 (bases 1 to 1584)
Isogai,T. and Yamamoto,J.
Direct Submission
Submitted (04-JUL-2002) Takao Isogai, RUI Project(HRI Team); 2-6-7
Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
NEO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing;
Research Association for Biotechnology (RAB); cDNA library

```

```

Construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'-6' end one pass sequencing; RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing; HRI and
RAB; annotation: HRI and RAB.
FEATURES
source
1.1584
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="SALGL1000107"
/tissue="salivary gland"
/clone_lib="SALGL1"
/Note="Cloning vector: pME18SF13"
89.997
CDS
/Note="unannotated protein product"
/codon_start=1
/protein_id="BAC03943.1"
/db_xref="GI:21751328"
/translation="MGDSEWLKLPVOKCKHKLKRLSGEALKIPKIDKSP
EMSKFGLIKKEVDSNANVYOLKGLAALVYENAHVAGTTGVSQVSKFNQK
AKAEIGETIKETIETIEKRAVEELKGLDNKPKIIVACIETTLKRALSEFSKII
ILKPIITKPLKPSREKAVDEAKLIVETIRWIRDALPPIQINISVOLKLEDEM
VLPFSAPRPFRFSQOLEAKLEQOOSAGDAGEGDDGDEVPQIDAVELLEAVEI
LSKLPRKFMRLRKQKGRKRPMSL"
526 a 298 c 331 g 429 t
BASE COUNT
ORIGIN
Alignment Scores:
Pred. No.: 1.02e-10 Length: 1584
Score: 178.50 Matches: 37
Percent Similarity: 63.77% Conservative: 7
Best Local Similarity: 53.62% Mismatches: 8
Query Match: 34.20% Indels: 17
DB: 9 Gaps: 2

US-09-830-244b-1 (1-95) x AK092678 (1-1584)
OY 27 LeuValSerLeuLysLysLysProPheHisLeuTyrCysValIleTyrIleProLeu 46
Db 1152 CTGATTCACTCGAAGAGAAAT-----TTTCCGTA-----1181
OY 47 ValProLysLeuIleIleLeuPheLeuAspIleAlaPheIleProLysSerLeuIleSer 66
Db 1182 -----GAATTGGAGATTCCGTTATGGGTATGGCCCTTATCAGC 1220
OY 67 GluPheGlnAsnAsnHisTyrThrHisAsnHisThrAsnHisAsnThrAsnAsnIleArg 86
Db 1221 CAGTTCCAGAACACCACTATACCCACACCATACCAACCACTACCAATATATACCT 1280
OY 87 PheAsnIleIleSerAsnCysArgThr 95
Db 1281 TTTAATATCATCACTACTGACGAGACA 1307

RESULT 7
HUMSTAT1A 4723 bp DNA linear PRI 13-JAN-1995
LOCUS HUMSTAT1A Homo sapiens statherin gene, exons 2-6.
DEFINITION HUMSTAT1A Homo sapiens statherin gene, exons 2-6.
ACCESSION M32639
VERSION M32639.1 GI:338504
KEYWORDS statherin.
SOURCE Human (individuals #563, #8136, and J.F.) fibroblast, cell line
#563, DNA, clones 1-3.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 4723)
Sabbatini,L.M., He,Y.Z. and Azen,E.A.
Structure and sequence determination of the gene encoding human
salivary statherin
Gene 89 (2), 245-251 (1990)
JOURNAL MEDLINE
PUBMED 90333623
2373369

```

LOCUS	141568 bp	DNA	linear	PRI 25-AUG-2000
DEFINITION	AC063956	Homio sapiens 4 BAC RP11-529K3 (Roswell Park Cancer Institute Human BAC library) complete sequence.		
ACCESSION	AC063956			
VERSION	AC063956.7	GI:9910030		
KEYWORDS	HTG.			
SOURCE	Homio sapiens.			
ORGANISM	Homio sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
AUTHORS	1 (bases 1 to 141568) Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alspirooks,S.L., Amarantunge,H.C., Are,J.R., Banks,T., Barbatale,J., Benton,J., Binagoe,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowe,S., Bruleva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Buretek,S.C., Burrell,K.L., Byrd,N.C., Cartron,T.F., Carter,H., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwalte,K.J., Draper,H., Dugan-rocha,S., Durbin,K.D., Earnhart,C., Edgar,D., Edwards,C.C., Ellias,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Gastara,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., He,X., Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B., Homsl,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Ioshikhes,I., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jollivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., Kling,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leil,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Iieu,C., Liu,J., Iliu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Matindale,A., Martinez,E., Massey,E., Mawhney,E., Mcclodo,M.P., Meedor,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Moalbadt,K., Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neel,D., Nelson,D., Newton,J., Newton,S., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokwkw,S., Oguh,M., Okumou,G., Orgunye,N., Oviedo,R., Pace,A., Payton,B., Peedy,T., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Qlles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoocharai,N., Sisson,I., Soederren,E., Sonake,T., Sparks,A., Stanley,H., Stone,K., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,H., Tang,H., Tansey,K., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,Y., Villalta,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S., Williams,G., Williamson,A., Wlezyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Kucherlapati,R., Naylor,S.L. and Gibbs,R.			
TITLE	Direct Submission			
JOURNAL	2 (bases 1 to 141568)			
REFERENCE	Unpublished			
AUTHORS	Worley,K.C.			
TITLE	Direct Submission			
JOURNAL	Submitted (22-APR-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA			
REFERENCE	3 (bases 1 to 141568)			
AUTHORS	Worley,K.C.			
TITLE	Direct Submission			
JOURNAL	Submitted (25-AUG-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA			
COMMENT	On Aug 25, 2000 this sequence version replaced at: 9765448			



OY 88 Asn1le1leSeracnysAargThr 95  
 Db 43077 AATATCATCACTACTGACAGACA 43100  
 RESULT 9  
 AC024676 161549 bp DNA linear HTG 16-MAR-2000  
 LOCUS  
 DEFINITION Homo sapiens chromosome 4 clone RP11-751H12 map 4, WORKING DRAFT  
 AC024676  
 AC024676.2 GI:7249372  
 VERSION HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
 KEYWORDS  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 161549)  
 Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
 Homo sapiens chromosome 4, clone RP11-751H12  
 Unpublished  
 2 (bases 1 to 161549)  
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
 Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,  
 Boguslavsky, L., Boukhalter, B., Brown, A., Burkett, G.,  
 Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,  
 Collamore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S.,  
 Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, M., Gage, D.,  
 Galagan, J., Gardyna, S., Glade, S., Goyette, M., Graham, L.,  
 Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,  
 Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,  
 Klein, J., Larocque, K., Lamazares, R., Landers, T., Lehoczy, J.,  
 Levine, R., Liu, C., Liu, G., Locke, K., MacDonald, P., Marquis, N.,  
 McCarthy, M., McEwan, P., McGuire, A., McKernan, K., McPheters, R.,  
 Melchior, J., Meneses, L., Milnova, T., Miranda, C., Mlenga, V., Morrow, J.,  
 Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,  
 O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N.,  
 Plesni, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,  
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,  
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
 Tassfay, S., Theodore, J., Tirrell, A., Travers, M., Triggillo, J.,  
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,  
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (01-MAR-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Mar 16, 2000 this sequence version replaced gi:7139892.  
 All repeats were identified using RepeatMasker:  
 Smit, A. P. A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

\* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 1 1060: contig of 1060 bp in length  
 \* 1061 1160: gap of 100 bp  
 \* 1161 2376: contig of 1216 bp in length  
 \* 2377 2476: gap of 100 bp  
 \* 2477 3989: contig of 1513 bp in length  
 \* 3990 4089: gap of 100 bp  
 \* 4090 5305: contig of 1216 bp in length  
 \* 5306 5405: gap of 100 bp  
 \* 5406 7418: contig of 2013 bp in length  
 \* 7419 7518: gap of 100 bp  
 \* 7519 9669: contig of 2151 bp in length  
 \* 9670 9769: gap of 100 bp  
 \* 9770 11064: contig of 1295 bp in length  
 \* 11065 11164: gap of 100 bp  
 \* 11165 14241: contig of 3077 bp in length  
 \* 14242 14341: gap of 100 bp  
 \* 14342 17832: contig of 3491 bp in length  
 \* 17833 17932: gap of 100 bp  
 \* 17933 22345: contig of 4413 bp in length  
 \* 22346 22445: gap of 100 bp  
 \* 22446 27054: contig of 4609 bp in length  
 \* 27055 27154: gap of 100 bp  
 \* 27155 31430: contig of 4276 bp in length  
 \* 31431 31530: gap of 100 bp  
 \* 31531 37872: contig of 6342 bp in length  
 \* 37873 37972: gap of 100 bp  
 \* 37973 44394: contig of 6422 bp in length  
 \* 44395 44494: gap of 100 bp  
 \* 44495 51121: contig of 6627 bp in length  
 \* 51122 51221: gap of 100 bp  
 \* 51222 58058: contig of 6837 bp in length  
 \* 58059 58158: gap of 100 bp  
 \* 58159 67135: contig of 8977 bp in length  
 \* 67136 67235: gap of 100 bp  
 \* 67236 75851: contig of 8616 bp in length  
 \* 75852 75951: gap of 100 bp  
 \* 75952 84245: contig of 8294 bp in length  
 \* 84246 84345: gap of 100 bp  
 \* 84346 95422: contig of 11077 bp in length  
 \* 95423 95522: gap of 100 bp  
 \* 95523 108795: contig of 13273 bp in length  
 \* 108796 108895: gap of 100 bp  
 \* 108896 135384: contig of 26489 bp in length  
 \* 135385 135484: gap of 100 bp  
 \* 135485 161549: contig of 26065 bp in length.  
 FEATURES  
 source  
 1..161549  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="4"  
 /map="4"  
 /clone="RP11-751H12"  
 /clone\_id="RP11-751H12 Human Male BAC"  
 1..1060  
 /note="assembly-fragment"  
 1161..2376  
 /note="assembly-fragment"  
 2477..3989  
 /note="assembly-fragment"  
 4090..5305  
 /note="assembly-fragment"  
 5406..7418  
 /note="assembly-fragment"  
 7519..9669  
 /note="assembly-fragment"  
 9770..11064  
 /note="assembly-fragment"  
 clone\_end:T7  
 vector\_side:left

TITLE  
 JOURNAL  
 COMMENT  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 Project Information  
 Center Project name: L7432  
 Center Clone name: 751\_H\_12  
 Summary Statistics  
 Sequencing vector: M13: M77815; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 148549 bases at least Q40  
 Consensus quality: 14910 bases at least Q30  
 Consensus quality: 157453 bases at least Q20  
 Insert size: 157000; agarose-fp  
 Insert size: 159349; sum-of-contigs  
 Quality coverage: 3.8 in Q20 bases; sum-of-contigs  
 Quality coverage: 3.8 in Q20 bases; sum-of-contigs  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 23 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as

```

misc_feature      11165..14241
                    /note="assembly-fragment"
misc_feature      14342..117832
                    /note="assembly-fragment"
misc_feature      17933..22345
                    /note="assembly-fragment"
misc_feature      22446..27054
                    /note="assembly-fragment"
misc_feature      27155..31430
                    /note="assembly-fragment"
misc_feature      31531..37872
                    /note="assembly-fragment"
misc_feature      37973..44394
                    /note="assembly-fragment"
misc_feature      44495..51121
                    /note="assembly-fragment"
misc_feature      51222..58058
                    /note="assembly-fragment"
misc_feature      58159..67135
                    /note="assembly-fragment"
misc_feature      67236..75851
                    /note="assembly-fragment"
misc_feature      75952..84245
                    /note="assembly-fragment"
misc_feature      84346..95422
                    /note="assembly-fragment"
misc_feature      clone.end.SP6
                    /note="assembly-fragment"
misc_feature      95523..108795
                    /note="assembly-fragment"
misc_feature      108896..135384
                    /note="assembly-fragment"
misc_feature      135485..161549
                    /note="assembly-fragment"
BASE COUNT      53276 a 27644 c 27367 g 51058 t 2204 others
ORIGIN

```

```

Alignment Scores:
Pred. No.:      1.38e-08      Length:      161549
Score:          178.00      Matches:      34
Percent Similarity: 77.08%      Conservative: 3
Best Local Similarity: 70.83%      Mismatches: 5
Query Match:     34.10%      Indels:      6
                        Gaps:      1

```

US-09-830-244B-1 (1-95) x AC024676 (1-161549)

```

QY      54 Phlebaspi1le1aphe-----1leProlyserleu1leSergln 67
          |||||      |||||      |||||      |||||      |||||
Db 125221 TTCTGCAATTTCTTCTCTGTTGTTACAGTATGCGTATGCGCTTATCAGCCAG 125280
          |||||      |||||      |||||      |||||      |||||
QY      68 Phcglanasn1styt1hr1h1asn1sthr1asn1sthr1asn1le1arpye 87
          |||||      |||||      |||||      |||||      |||||
Db 125281 TTCCAGAACACACATATCCCAACATACCAACATACCAACATATACCTTTT 125340
          |||||      |||||      |||||      |||||      |||||
QY      88 Asn1le1leSern1ncysar1gthr 95
          |||||      |||||      |||||      |||||      |||||
Db 125341 AATATCATCATGTAATCGACAGACA 125364
          |||||      |||||      |||||      |||||      |||||

```

```

RESULT 10
AC021462/c
LOCUS      AC021462      153023 bp      DNA      linear      HTG 03-APR-2000
DEFINITION      Homo sapiens clone RP11-276C1, WORKING DRAFT SEQUENCE, 21 unordered
pieces.
AC021462
VERSION      AC021462.3      GI:7387343
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 153023)
AUTHORS      Birren,B., Linton,L., Nusbaum,C. and Lander,E.

```

# TITLE JOURNAL REFERENCE AUTHORS

Homo sapiens, clone RP11-276C1

Unpublished  
2 (bases 1 to 153023)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Anderson,S., Baldwin,J., Barna,N., Beckel,R., Beda,F.,  
Boguslavsky,L., Bouckgalter,B., Brown,A., Burkett,G., Castle,A.,  
Chapel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,  
Dearliano,K., Dewar,K., Domino,M., Doyle,W., Fenesior,D.,  
Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,  
Gardyna,S., Grant,G., Hagos,B., Heatford,A., Horton,L.,  
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,  
Landers,T., Lechoczky,J., Levine,R., Lien,C., Liu,G., Locke,K.,  
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,  
McPheters,R., Meldrum,J., Menessier,L., Morrill,T., Naylor,J.,  
Norman,C.H., O'Connor,T., O'Donnell,P., Ollivar,T.M., Peterson,K.,  
Pierre,N., Pisanic,C., Pollara,V., Raymond,C., Riley,R., Rochman,D.,  
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,  
Stojanovic,N., Subramanian,A., Talamas,J., Teafaye,S., Theodore,J.,  
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,  
Zimmer,A. and Zody,M.

## TITLE JOURNAL

### COMMENT

Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Apr 3, 2000 this sequence version replaced gi:7230200.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence\_submissions@genome.wi.mit.edu

Project Information

Center project name: 15201

Center clone name: 276C\_1

Summary Statistics

Sequencing vector: M13; M7815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 141815 bases at least Q40  
Consensus quality: 147368 bases at least Q30  
Consensus quality: 149481 bases at least Q20  
Insert size: 157000; agarose-fp  
Insert size: 151023; sum-of-ctrls  
Quality coverage: 4.3 in Q20 bases; agarose-fp  
Quality coverage: 4.5 in Q20 bases; sum-of-ctrls

NOTE: This is a 'working draft' sequence. It currently  
consists of 21 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.

1 1015: contig of 1015 bp in length  
1016 1115: gap of 100 bp  
1116 2276: contig of 1161 bp in length  
2277 2376: gap of 100 bp  
2377 4009: contig of 1633 bp in length  
4010 4109: gap of 100 bp  
4110 5737: contig of 1628 bp in length  
5738 5837: gap of 100 bp  
5838 6939: contig of 1102 bp in length  
6940 7039: gap of 100 bp  
7040 9504: contig of 2465 bp in length  
9505 9604: gap of 100 bp  
9605 11734: contig of 2130 bp in length  
11735 11834: gap of 100 bp  
11835 14111: contig of 2277 bp in length  
14112 14211: gap of 100 bp  
14212 16973: contig of 2762 bp in length  
16974 17073: gap of 100 bp



	BASE COUNT	41494	a	35192	c	34573	g	39756	t	2008	others
	ORIGIN										
	Alignment Scores:										
	Pred. No.:	9.	39e-05	Length:	153023						
	Score:	143.50	Matches:	31							
	Percent Similarity:	53.3%	Conservative:	9							
	Best Local Similarity:	41.3%	Mismatches:	13							
	Query Match:	27.49%	Indels:	22							
	DB:	2	Gaps:	1							
US-09-830-244B-1 (1-95) x AC021462 (1-153023)											
QY	1 MetcTrPheHISLysValcIyArGLysGlnHisphelysValThrPhetrgIurrrasp	20									
Db	13988 ATGTGGTTTCATCCACAGTGGGAGAAAGCAGCAATTATAAATACTTTTGGAGACTGA	13929									
QY	21 LeuSeAsAspIlysthr-----LeuValserLeuylsylysylsySprophenIsleutyr	26									
Db	13928 TTGAGTAATATTAACAACCTTGACTCTTTCGCATTAATTAATTAATTAATTAATTAACAA	13869									
QY	27 -----LeuValserLeuylsylysylsySprophenIsleutyr	39									
Db	13868 CAACTATTGAATGGCGCACCTGACTAGATGAGAGAAGAGCAATTTTTGCGATTTC	13900									

0Y	40	-CysValIleTyrIleProLeuValProTyrIleIleLeu	53		
Db	13808	TTGCCTAGTTTCCCTTATAAGCACCACTAAGTTAATAGCTCTG	13766		
RESULT 11					
	LOCUS	AX281311	9881 bp	DNA	linear
	DEFINITION	Sequence 53 from Patent WO0177164.			PAT 02-NOV-2001
	ACCESSION	AX281311			
	VERSION	AX281311.1	GI:1608566		
	KEYWORDS				
	SOURCE	synthetic construct.			
	ORGANISM	synthetic construct			
		artificial sequences.			
	REFERENCE	1			
	AUTHORS	Olek, A., Piepenbrock, C. and Berlin, K.			
	TITLE	Diagnosis of diseases associated with apoptosis			
	JOURNAL	Patent: WO 0177164-A 53 18-Oct-2001;			
		Epigenomics AG (DE)			
FEATURES					
	source	Location/Qualifiers			
		1..9881			
		/organism="synthetic construct"			
		/db_xref="taxon:32630"			
		/note="Chemically treated genomic DNA (Homo sapiens)"			
	BASE COUNT	2606 a 196 c 2308 g 4771 t			
	ORIGIN				
Alignment Scores:					
	Pred. No.:	0.00574	Length:	9881	
	Score:	116.50	Matches:	28	
	Percent Similarity:	51.32%	Conservative:	11	
	Best Local Similarity:	36.84%	Mismatches:	15	
	Query Match:	22.32%	Indels:	22	
	DB:	6	Gaps:	1	
US-09-830-244B-1 (1-95) x AX281311 (1-9881)					
0Y	1	MettirPheHisLysValGlyArgLysGlnHisPheLysValIthrPheTrpGluThrAsp	20		
Db	5782	ATGTGGTTTCTTTTAAAGCGGCAAGAACTAGTAATTAAATAATTTTGGGACATTGAA	5841		:::
0Y	21	LeuSerAsnAsnLysThrLeuValSerLeu	30		
Db	5842	TTTGAGTAATAATAAATTAGTTTTCGTTAATAATAATAATAATAATAATAATAA	5901		
0Y	31	-----LysLysLysLysProPheHisLeu	Ty 39		



Db 131844 ATGTGATTCACACGCTTTGGGCAAAATGAAGTTCAGGACTTAAGCAGTTAT-GAAGCA 131902  
 OY 20 AspleusSerAsnAsnlyThrLeuVal---SerLeuIys-----31  
 Db 131903 GACGTGGACATCATATTAAGATTATTAATCTTATTATAGTATAAGAAACAGCATACCT 131962  
 OY 32 -----LysLysIysProPheHis-----Leu 38  
 Db 131963 GAGACATGATGGTATGGCGCTTCACAAAGACACACCTATCTCAGTACTGTTTGTGTTTG 132022  
 OY 39 TyrCysValIleIleTyrIleProLeuValProLysLeuIleIleLeuPheLeuAspIleAla 58  
 Db 132023 TTTGTTGTTTTCACACATTAATCAATATACAGATTTTGGTAGCGCTTGTACTATATTT 132082  
 OY 59 PheIleProLysSerIleIleSerGlnPheGlnAsnHis-----TyrThrHis 75  
 Db 132083 -----AAAGGTTCCTTAAGACACACTTAATATGGACACACACACACACACACACACAC 132133  
 OY 76 AsnHisThrAsnHisAsnThrAsnAsnIle 85  
 Db 132134 ACACACACACACACACACACACACACAGTGT 132163

RESULT 13  
 AC100542/C 55219 bp DNA linear HNG 22-NOV-2001  
 LOCUS Mus musculus clone RP23-152620, LOW-PASS SEQUENCE SAMPLING.  
 DEFINITION AC100542  
 AC100542.1 GI:17047908  
 VERSION HNG: HNGS\_PHASD0.  
 KEYWORDS Mus musculus.  
 SOURCE Mus musculus.  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
 TITLE Mus musculus, clone RP23-152620  
 JOURNAL Unpublished  
 AUTHORS 2 (bases 1 to 55219)  
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boucknight, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choquel, Y., DeAngelis, M., Collins, S., Collymore, A., Cook, A., Cooke, P., DeAngelis, M., Dewar, K., Diaz, J. S., Dodge, S., Farcis, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Glend, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Lacombe, K., Lamaszates, R., Landers, T., Lechoczy, V., Levine, R., Liu, G., Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheters, R., Meldrum, J., Menues, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollard, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliou, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE Direct Submission  
 JOURNAL Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 COMMENT All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/MIT Center for Genome Research  
 Center code: WITR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: L15664

Center clone name: 152\_G\_20

\* NOTE: This record contains 69 individual  
 \* sequencing reads that have not been assembled into  
 \* contigs. Runs of N are used to separate the reads  
 \* and the order in which they appear is completely  
 \* arbitrary. Low-pass sequence sampling is useful for  
 \* identifying clones that may be gene-rich and allows  
 \* overlap relationships among clones to be deduced.  
 \* However, it should not be assumed that this clone  
 \* will be sequenced to completion. In the event that  
 \* the record is updated, the accession number will  
 \* be preserved.

1 729: contig of 729 bp in length  
 \* 730 829: gap of 100 bp  
 \* 830 1532: contig of 703 bp in length  
 \* 1533 1632: gap of 100 bp  
 \* 1633 2340: contig of 708 bp in length  
 \* 2341 2440: gap of 100 bp  
 \* 2441 3151: contig of 711 bp in length  
 \* 3152 3251: gap of 100 bp  
 \* 3252 3940: contig of 689 bp in length  
 \* 3941 4040: gap of 100 bp  
 \* 4041 4731: contig of 691 bp in length  
 \* 4732 4831: gap of 100 bp  
 \* 4832 5517: contig of 686 bp in length  
 \* 5518 5617: gap of 100 bp  
 \* 5618 6321: contig of 704 bp in length  
 \* 6322 6421: gap of 100 bp  
 \* 6422 7135: contig of 714 bp in length  
 \* 7136 7235: gap of 100 bp  
 \* 7236 7967: contig of 732 bp in length  
 \* 7968 8067: gap of 100 bp  
 \* 8068 8795: contig of 728 bp in length  
 \* 8796 8895: gap of 100 bp  
 \* 8896 9582: contig of 687 bp in length  
 \* 9583 9682: gap of 100 bp  
 \* 9683 10388: contig of 706 bp in length  
 \* 10389 10488: gap of 100 bp  
 \* 10489 11188: contig of 700 bp in length  
 \* 11189 11288: gap of 100 bp  
 \* 11289 11975: contig of 687 bp in length  
 \* 11976 12075: gap of 100 bp  
 \* 12076 12765: contig of 690 bp in length  
 \* 12766 12865: gap of 100 bp  
 \* 12866 13566: contig of 701 bp in length  
 \* 13567 13666: gap of 100 bp  
 \* 13667 14362: contig of 696 bp in length  
 \* 14363 14462: gap of 100 bp  
 \* 14463 15163: contig of 701 bp in length  
 \* 15164 15263: gap of 100 bp  
 \* 15264 15958: contig of 695 bp in length  
 \* 15959 16058: gap of 100 bp  
 \* 16059 16758: contig of 700 bp in length  
 \* 16759 16858: gap of 100 bp  
 \* 16859 17561: contig of 703 bp in length  
 \* 17562 17661: gap of 100 bp  
 \* 17662 18379: contig of 718 bp in length  
 \* 18380 18479: gap of 100 bp  
 \* 18480 19153: contig of 674 bp in length  
 \* 19154 19253: gap of 100 bp  
 \* 19254 19950: contig of 697 bp in length  
 \* 19951 20050: gap of 100 bp  
 \* 20051 20739: contig of 689 bp in length  
 \* 20740 20839: gap of 100 bp  
 \* 20840 21538: contig of 699 bp in length  
 \* 21539 21638: gap of 100 bp  
 \* 21639 22336: contig of 698 bp in length  
 \* 22337 22436: gap of 100 bp  
 \* 22437 23145: contig of 709 bp in length  
 \* 23146 23245: gap of 100 bp  
 \* 23246 23957: contig of 712 bp in length  
 \* 23958 24057: gap of 100 bp

24058	24788:	contlg of 731 bp	in length
24789	24888:	gap of 100 bp	
24889	25597:	contlg of 709 bp	in length
25598	25697:	gap of 100 bp	
25698	26400:	contlg of 703 bp	in length
26401	26500:	gap of 100 bp	
26501	27188:	contlg of 688 bp	in length
27189	27288:	gap of 100 bp	
27289	27990:	contlg of 702 bp	in length
27991	28090:	gap of 100 bp	
28091	28774:	contlg of 684 bp	in length
28775	28874:	gap of 100 bp	
28875	29591:	contlg of 717 bp	in length
29592	29691:	gap of 100 bp	
29692	30427:	contlg of 736 bp	in length
30428	30527:	gap of 100 bp	
30528	31223:	contlg of 698 bp	in length
31226	31325:	gap of 100 bp	
31326	32039:	contlg of 714 bp	in length
32040	32139:	gap of 100 bp	
32140	33841:	contlg of 702 bp	in length
32982	32941:	gap of 100 bp	
32982	33624:	contlg of 683 bp	in length
33625	33724:	gap of 100 bp	
33725	34415:	contlg of 691 bp	in length
34416	34515:	gap of 100 bp	
34516	35221:	contlg of 706 bp	in length
35222	35331:	gap of 100 bp	
35322	36031:	contlg of 710 bp	in length
36032	36131:	gap of 100 bp	
36132	36855:	contlg of 727 bp	in length
36859	36958:	gap of 100 bp	
36959	37684:	contlg of 726 bp	in length
37685	37784:	gap of 100 bp	
37785	38481:	contlg of 697 bp	in length
38482	38581:	gap of 100 bp	
38582	39282:	contlg of 701 bp	in length
39283	39382:	gap of 100 bp	
39383	40077:	contlg of 685 bp	in length
40078	40177:	gap of 100 bp	
40178	40845:	contlg of 668 bp	in length
40846	40945:	gap of 100 bp	
40946	41633:	contlg of 678 bp	in length
41624	41723:	gap of 100 bp	
41724	42433:	contlg of 710 bp	in length
42534	42533:	gap of 100 bp	
42534	43244:	contlg of 711 bp	in length
43245	43344:	gap of 100 bp	
43345	44050:	contlg of 706 bp	in length
44051	44150:	gap of 100 bp	
44151	44647:	contlg of 697 bp	in length
44648	44947:	gap of 100 bp	
44948	45656:	contlg of 709 bp	in length
45657	45756:	gap of 100 bp	
45757	46470:	contlg of 714 bp	in length
46471	46570:	gap of 100 bp	
46571	47262:	contlg of 692 bp	in length
47263	47362:	gap of 100 bp	
47363	48041:	contlg of 679 bp	in length
48042	48141:	gap of 100 bp	
48142	48643:	contlg of 702 bp	in length
48644	48943:	gap of 100 bp	
48944	49638:	contlg of 695 bp	in length
49639	49738:	gap of 100 bp	
49739	50416:	contlg of 678 bp	in length
50417	50516:	gap of 100 bp	
50517	51244:	contlg of 728 bp	in length
51245	51344:	gap of 100 bp	
51345	52051:	contlg of 707 bp	in length
52052	52151:	gap of 100 bp	
52152	52843:	contlg of 692 bp	in length
52844	52943:	gap of 100 bp	
52944	53621:	contlg of 678 bp	in length

FEATURES	*	*	*	*
53622	53721:	gap of	100 bp	
53722	54421:	contig of	700 bp	in length
54422	54521:	gap of	100 bp	
54522	55219:	contig of	698 bp	in length
	Location/Qualifiers			





GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OW protein - nucleic search, using frame\_plus\_p2n model

Run on: April 14, 2003, 23:05:11; Search time 225 Seconds

(Without alignments)  
950.844 Million cell updates/sec

Title: US-09-830-244B-1

Perfect score: 522  
Sequence: 1 MWFHKVGRKHFKFTFWETD.....NHTNHTNNTNIFNIIISCR 95

Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 2185239 segs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-MODEL-frame+ p2n.model -DEV-xip  
-O/cgn2.1/USPTO.spool/US09830244/runat\_10042003\_090015\_27856/app\_query.fasta\_1.263  
-DB-N\_Geneseq\_101002 -OPT-fastlap -SDFE1-rng -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS-bits -START=1 -END=1 -MATRIX-blosunb2 -TRANS-human40.cdl  
-LIST=45 -DOCALLIGN=200 -THR\_SCORE-pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE-LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MNTLEN=0 -MAXLEN=2000000000  
-USRR-US09830244.cgn.1.1.396.0runat.10042003.090015.27856 -NCP=6 -ICPU=3  
-NO\_XLPHY -NO\_MMAP -LARGEJURY -NEG\_SCORES=0 -MAIT -LONCLOG -DEV-TIMEOUT=120  
-MARN\_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: N\_Geneseq\_101002:\*

1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Length	DB	ID	Description
1	522	100.0	1331	21	AAA48963	Human lysine-rich
2	200	38.3	221	21	AAA42355	Human secreted exp
3	178.5	34.2	552	21	AAA48964	Human statherin DN
4	178.5	34.2	2121	21	AAC76655	Human ORFX ORF220
5	136	26.1	390	20	AA191635	Human polynucleoti
6	121.5	23.3	516	20	AAV89555	EST clone CP289.
7	117	22.4	126	16	AA124320	Human gene signatu
8	116.5	22.3	9881	24	AB154353	Chemically treated
9	113	21.6	60	24	ABN43273	Human spliced tran
10	101	19.3	6070	24	AB199219	Chemically treated
11	101	19.3	6070	24	AB149310	Human polynucleoti
12	101	19.3	6070	24	AB132241	Human immune syste
13	92	17.6	9021	22	AA546326	Tumour suppressor
14	91.5	17.5	7038	22	AA545466	Chemically pretrea
15	91.5	17.5	7038	24	AB134097	Human immune syste
16	91.5	17.5	7038	24	ABR28426	DNA transcription
17	91.5	17.5	9881	24	AB154354	Chemically treated
18	89.5	17.1	8246	24	AB152202	Human immune syste
19	87.5	16.8	32205	22	ABA08217	Human ovarian and
20	87.5	16.8	32205	22	AA107525	Human reproductive
21	87.5	16.8	45186	22	AA107525	Wild-type human CT
22	86	16.5	6754	24	AB170346	Human gene regulat
23	86	16.5	6754	24	AA561305	Human immune syste
24	85	16.3	6151	24	AB132622	Human secreted pro
25	83.5	16.0	1136	21	AAC59338	Human immune syste
26	83.5	16.0	6535	24	AB132936	Human immune syste
27	83.5	16.0	12571	21	AA235272	Soybean retroelme
28	82.5	16.0	16545	24	AB132051	Human immune syste
29	82.5	15.8	1791	23	AB105983	Drosophila melanog
30	82.5	15.8	8889	23	AB105982	Human immune syste
31	82	15.7	8889	24	AB132211	Tumour suppressor
32	81.5	15.6	6031	22	AA546522	Human immune syste
33	81.5	15.6	13376	24	AB132582	Human polynucleoti
34	80	15.3	556	22	AA191773	Human polynucleoti
35	80	15.3	5666	24	AB133772	Human immune syste
36	80	15.3	6059	24	AB133481	Human immune syste
37	80	15.3	6350	24	AB170500	Human immune syste
38	80	15.3	6350	24	AB134629	Chemically treated
39	79.5	15.2	13814	24	AB133193	Human immune syste
40	79	15.1	8996	22	AA545503	Human metastasis a
41	79	15.1	8996	24	ABR28435	DNA transcription
42	78.5	15.0	10460	23	AB103520	Drosophila melanog
43	78.5	15.0	10514	23	AB120932	Drosophila melanog
44	78.5	15.0	17722	23	AB103508	Drosophila melanog
45	78.5	15.0	30037	22	ABA20038	Human nervous syst

## ALIGNMENTS

RESULT 1  
AAA48963  
ID AAA48963 standard; cDNA; 1331 BP.

AC AAA48963;  
XX  
DT 06-OCT-2000 (first entry)

DE Human lysine-rich statherin cDNA from Incyte clone 2820214.

XX Lysine-rich statherin protein; LRSP; acidic phosphoprotein; human;  
KW precipitation inhibitor; autoimmune; inflammatory disorder; AIDS;  
KW asthma; allergy; diabetes mellitus; fungal; bacterial infection;  
KW cancer; leukemia; adenocarcinoma; melanoma; ss.

OS Homo sapiens.

XX Key Location/Qualifiers



FT CDS 871..1158  
 FT /\*tag= a  
 FT /product= Lysine-rich statherin  
 PN MO200024779-A1.  
 XX 04-MAY-2000.  
 XX 22-OCT-1999; 99WO-US24046.  
 XX 23-OCT-1998; 98US-0155209.  
 XX (INCY-) INCYTE PHARM INC.  
 XX Tang YT, Corley NC, Guejler KU, Patterson C;  
 DR WPI: 2000-350699/30.  
 DR P-PSDB: AAY94526.  
 XX  
 PT Purified polypeptide used for treating or preventing a disorder  
 PT characterized by expression or activity of lysine-rich statherin  
 PT proteins -  
 XX  
 PS Claim 3; Page 70; 75pp: English.  
 XX  
 CC The present sequence is human lysine-rich statherin protein (LRSP)  
 CC cDNA from Incyte clone 2620214. This sequence was identified  
 CC through analysis of a cDNA library of breast tumour tissue  
 CC (BRSTW0114). The LRSP sequence was found to have homology with  
 CC human statherin (AAY94527) and human basic histidine-rich protein  
 CC (AAY94528). Human statherin is a phosphoprotein that acts as an  
 CC inhibitor of precipitation of calcium phosphate salts in the oral  
 CC cavity. The LRSP polypeptide and its antagonists may be useful for  
 CC treating or preventing disorders associated with the activity of  
 CC LRSP. Such disorders include autoimmune/inflammatory disorders (for  
 CC example AIDS, allergies, asthma, diabetes mellitus), bacterial and  
 CC fungal infection and cancers (such as leukemia, adenocarcinoma,  
 CC melanoma). Antibodies to LRSP may be useful for diagnosis of  
 CC the above disorders.  
 CC  
 SO Sequence 1331 BP; 397 A; 258 C; 228 G; 447 T; 1 other;  
 Alignment Scores:  
 Pred. No.: 1.55e-53 Length: 1331  
 Score: 522.00 Matches: 95  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 21 Gaps: 0  
 US-09-830-244B-1 (1-95) x AAA48963 (1-1331)  
 QY 1 MetTTPhehIstysValiGlyAglvGlnhIshPhelYsValThrPhetPglurhAsp 20  
 DB 871 ATGTGGTTTCATAAAGTGGGAGAGAAAAACACATTTTAAAGTAAGTCTTTGGAGACTGAT 930  
 QY 21 LeuSerAsnAsnIstThrLeuValSerLeuIstYsValYsPhehIstYsCys 40  
 DB 931 TTGAGTAATTAATAAACTCGCTCGCTCCCTTAAGAAAAAAACCTTCCACCTTACTGT 990  
 QY 41 ValIleTyTleProLeuValProIstYsValIleLeuPhehIstYsPhehIleAaPhele 60  
 DB 991 GTCATTAAATCCCTTACCTTCCAAAGTAAATTAATCTTATTCCTGCAATTCGTTTAA 1050  
 QY 61 ProIstSerLeuIstSerGlnPhehIstAsnAsnIstYsThhIstAsnIstAsnIst 80  
 DB 1051 CCAAGAGGCTTATCCAGGCAAGTCCAGAAACCACTATACGACAAACCATCAACAC 1110  
 QY 81 AsnThAsnAsnIleAtgPheAsnIleIstSerAsnCysArgThr 95  
 DB 1111 AATACCAACAATATACGTTTAAATATCATCACTGACAGAGACA 1155  
 RESULT 2

AAA42355  
 ID AAA42355 standard; cDNA: 221 BP.  
 XX  
 AC AAA42355;  
 XX  
 DT 21-AUG-2000 (first entry)  
 XX  
 DE Human secreted expressed sequence tag SEQ ID NO:1095.  
 XX  
 KW Human; mouse; xenopus; rat; secreted expressed sequence tag; SEST;  
 KW expressed sequence tag; EST; probe; chemotactic; proliferative;  
 KW immunomodulatory; haematopoietic; chemokine; analgesic; haemostatic;  
 KW thrombolytic; antiinflammatory; cyostatic; antibacterial; antifungal;  
 KW antiviral; antidiabetic; antiasclerotic; analgesic; antiparkinsonian;  
 KW antitumor; osteoprotective; neuroprotective; neurotropic; antiproliferative;  
 KW cerebroprotective; anticonvulsant; antidepressant; gene therapy;  
 KW vaccine; autoimmune disorder; multiple sclerosis; allergic condition;  
 KW insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;  
 KW lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;  
 KW central nervous system disorder; Huntington's disease; stroke;  
 KW Parkinson's disease; Huntington's disease; coagulation disorder;  
 KW haemophilia; thrombosis; inflammatory disorder; Crohn's disease;  
 KW tumour; infection; depression; psoriasis; ss.  
 KW  
 XX Homo sapiens.  
 OS  
 XX  
 PN WO200021990-A1.  
 XX  
 PD 20-APR-2000.  
 XX  
 PF 15-OCT-1999; 99WO-US24205.  
 XX  
 PR 15-OCT-1998; 98US-0104435.  
 XX  
 PA (GENM) GENNETICS INST INC.  
 XX  
 PI Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;  
 PI Merberg D, Treacy M;  
 PI  
 DR WPI: 2000-317937/27.  
 XX  
 PT Isolated polynucleotides, and encoded proteins, comprising secreted  
 PT expressed sequence tags (SESTs), useful for treating various disorders  
 PT such as autoimmune, infectious, and central nervous system disorders -  
 XX  
 PS Claim 1; Page 394; 618pp: English.  
 XX  
 CC AAA41261 to AAA43419 represent specifically claimed secreted expressed  
 CC sequence tags (SESTs), isolated from human, mouse, xenopus and rat  
 CC tissue sources. The SESTs can have a range of activities depending on  
 CC the tissues they were isolated from. The activities include:  
 CC chemotactic; proliferative; immunomodulatory; haematopoietic;  
 CC chemokine; analgesic; haemostatic; thrombolytic; antiinflammatory;  
 CC cytostatic; antibacterial; antifungal; antiviral; antidiabetic;  
 CC antiasclerotic; analgesic; antitumor; antiparkinsonian;  
 CC neurotropic; antiparkinsonian; antiproliferative; neuroprotective;  
 CC anticonvulsant; and antidepressant. The SESTs can be used for gene  
 CC therapy and in vaccines. The SESTs are useful as probes for the  
 CC identification and isolation of full-length cDNAs and genomic DNA  
 CC molecules which correspond to the SESTs. Proteins encoded by the SESTs  
 CC are useful in assays for determining biological activity and raising  
 CC antibodies. They may be useful for treatment of autoimmune disorders  
 CC (multiple sclerosis, insulin dependent diabetes), allergic conditions  
 CC (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,  
 CC osteoporosis, osteoarthritis, central nervous system disorders  
 CC (Alzheimer's, Huntington's, Parkinson's disease, stroke), coagulation  
 CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's  
 CC disease), tumours, bacterial, fungal or viral infections, depression and  
 CC psoriasis. AAA43420 to AAA43425 represent linker variants which are given  
 CC in the exemplification of the present invention.  
 XX  
 SO Sequence 221 BP; 72 A; 41 C; 44 G; 64 T; 0 other;



## Alignment Scores:

Pred. No.: 1 46e-15  
 Score: 200.00  
 Percent Similarity: 100.00%  
 Best Local Similarity: 100.00%  
 Query Match: 38.31%  
 DB: 21  
 Gaps: 0

US-09-830-244B-1 (1-95) x AAA48964 (1-221)

QY 1 MetTRPHEHISLysValGlyArgGlyGlnHisPhelYsValThrPheTrpGlnThrAsp 20  
 DB 106 ATGTGGTTTCATAAAGTGGGAGAAACAGCATTTTAAGTAACTTTTGGGAGACTGAT 165

QY 21 LeuSerAsnAsnLysThrLeuValSerLeuLysLysLysProPhe 36  
 DB 166 TTGAGTAAATATATAAACCTGTGCTCCCTTAAAGAAAAAACCCCTTC 213

## RESULT 3

AAA48964  
 ID AAA48964 standard; DNA; 552 bp.

AC AAA48964;

DT 06-OCT-2000 (first entry)

DE Human statherin DNA.

XX Lysine-rich statherin protein; LRSP; acidic phosphoprotein; human;  
 KW precipitation inhibitor; autoimmune; inflammatory disorder; AIDS;  
 KW asthma; allergy; diabetes mellitus; fungal; bacterial infection;  
 KW cancer; leukemia; adenocarcinoma; melanoma; ds.

XX Homo sapiens.

OS Key Location/Qualifiers

FT CDS 73..261

FT /product= Statherin

PN MO200024779-A1.

PD 04-MAY-2000.

PF 22-OCT-1999; 99MO-US24046.

PR 23-OCT-1998; 98US-0155209.

PA (INCY-) INCYTE PHARM INC.

PI Tang YF, Corley NC, Guegler KJ, Patterson C;

DR WPI: 2000-350699/30.

DR P-PSDB; AAY94527.

PT Purified polypeptide used for treating or preventing a disorder  
 PT characterized by expression or activity of lysine-rich statherin  
 PT proteins

PS Disclosure; Page 71; 75pp; English.

XX The present invention relates to human lysine-rich statherin protein  
 CC (LRSP)(AAY94526). The cDNA sequence encoding this protein was identified  
 CC through analysis of a cDNA library of breast tumour tissue (BRSTN014).  
 CC The LRSP sequence was found to have homology with the DNA of human  
 CC statherin (the present sequence) and human basic histidine-rich protein  
 CC (AAY94528). Human statherin is a phosphoprotein that acts as an  
 CC inhibitor of precipitation of calcium phosphate salts in the oral cavity.  
 CC The LRSP polypeptide and its antagonists may be useful for treating or  
 CC preventing disorders associated with the activity of LRSP. Such  
 CC disorders include autoimmune/inflammatory disorders (for example AIDS,  
 CC allergies, asthma, diabetes mellitus), bacterial and fungal infection  
 CC and cancers (such as leukemia, adenocarcinoma, melanoma). Antibodies to

CC LRSP may be useful for diagnosis of the above disorders.

SQ Sequence 552 bp; 172 A; 122 C; 78 G; 180 T; 0 other;

## Alignment Scores:

Pred. No.: 1 84e-12  
 Score: 178.50  
 Percent Similarity: 63.77%  
 Best Local Similarity: 53.62%  
 Query Match: 34.20%  
 DB: 21  
 Gaps: 2

US-09-830-244B-1 (1-95) x AAA48964 (1-552)

QY 27 LeuValSerLeuLysLysLysLysProPheHisLeuTyrCysValIleTyrIleProLeu 46  
 DB 128 CTGATTCATCTGAGAGAAAT-----TTTGGCTA----- 157

QY 47 ValProLysLeuIleIleLeuPheLeuAspIleAlaPheIleProLysSerIleIleSer 66  
 DB 158 -----GATTCGAGACTGCTGTTATGCGTATGCGCCCTTATCAGC 196

QY 67 GlnPheGlnAsnAsnHisTyrThrHisAsnHisThrAsnHisAsnThrAsnAsnIleArg 86  
 DB 197 CAGTTCAGAGACACACACATATACCCACAAACCATACCAACCAATATACATATACCT 256

QY 87 PheAsnIleIleSerAsnCysArgThr 95  
 DB 257 TTTAATATCATCATGATACCTGACGAGCA 283

## RESULT 4

AAC76665  
 ID AAC76665 standard; cDNA; 2121 bp.

AC AAC76665;

DT 08-FEB-2001 (first entry)

DE Human ORF2220 polynucleotide sequence SEQ ID NO:4439.

XX Human: open reading frame; ORF; detection; cytostatic; hepatotropic;  
 KW vulnery; antiproliferative; antiparkinsonian; neurotrophic; neuroprotective;  
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
 KW antiviral; antibacterial; antifungal; antineumatic; antithyroid;  
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;  
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KW cholesterol ester storage; systemic lupus erythematosus; infection;  
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 KW bone damage; cartilage damage; antinflammatory disease; coagulation;  
 KW thrombosis; contraceptive; ss.

XX Homo sapiens.

PN MO200058473-A2.

PD 05-OCT-2000.

PF 31-MAR-2000; 2000MO-US08621.

PR 31-MAR-1999; 99US-0127607.

PR 02-APR-1999; 99US-0127636.

PR 05-APR-1999; 99US-0127728.

PR 30-MAR-2000; 2000US-0540763.

PA (CURA-) CDRAGEN CORP.

PI Shinkets RA, Leach M;

DR WPI: 2000-602362/57.



```

XX OS Homo sapiens.
XX PN WO9845436-A2.
XX PD 15-OCT-1998.
XX PF 10-APR-1998; 98WO-US06955.
XX PR 10-APR-1997; 97US-0838821.
XX PA (GEM) GENETICS INST INC.
XX PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
XX PI Racie LA, Spaulding V, Treacy M;
XX DR WPI; 1999-070077/06.
XX
XX New polynucleotides encoding human secreted proteins - derived from
XX e.g. human blood, kidney, foetal lung, placenta, testes, brain,
XX ovary, pituitary, retina and colon cDNA libraries.
XX
XX Claim 1; Page 245; 618pp; English.
XX
XX The present sequence represents a human expressed sequence tag (EST).
XX The polynucleotide, which is a secreted EST, and the encoded protein
XX are predicted to have useful biological activities which would make
XX them suitable for treating, preventing or ameliorating medical
XX conditions in humans and animals, although no supporting data is
XX given. Suggested activities include nutritional activity, immune
XX stimulating or suppressing activity, haematopoiesis regulating
XX activity, tissue growth activity, activin/inhibin activity,
XX chemotactic/chemokinetic activity, haemostatic and thrombolytic
XX activity, receptor/ligand activity, anti-inflammatory activity,
XX adherin/tumour invasion suppressor activity, tumour inhibition
XX activity. The polynucleotide may also be useful for gene therapy.
XX
XX Sequence 516 BP; 156 A; 181 C; 97 G; 82 T; 0 other;

Alignment Scores:
Pred. No.: 1.33e-05 Length: 516
Score: 121.50 Matches: 28
Percent Similarity: 61.02% Conservative: 8
Best Local Similarity: 47.46% Mismatches: 16
Query Match: 23.28% Indels: 7
DB: 20 Gaps: 2

US-09-830-244B-1 (1-95) x AAV89525 (1-516)
QY 30 LeuLysLysLysLysProPheHisLeuTyxCysValIleTyrllePro----- 45
Db 41 TTGAGTAAAGAGAACCCAGCCAACTATGAGTTCCCTGTGCTTTCCTTCATCTTGCTC 100
QY 46 ---LeuValProLysLeuIleLeuPheLeuAspIleAlaPheIleProLysSerLeu 64
Db 101 TCATGGTTTCATGATGGAGCTGATCATCTG-----AAGACTAGGTATGGCCCTT 154
QY 65 IleserGlnPheGlnAsnHisTyrThrHisAsnHisThrAsnHisAsnThrAsn 83
Db 155 ATCAGCAGGTTCGACACACACCACTATACCCACACCAATACCAACCAATACCAAC 211
RESULT 7
AAV24320
ID AAT24320 standard; cDNA to mRNA; 126 BP.
XX AC AAT24320;
XX DT 22-SEP-1996 (first entry)
XX DE Human gene signature HUMGS06347.
XX
XX Gene signature; messenger RNA; mRNA; relative abundance; frequency;
XX human; cloning; mapping; non-biased library; diagnosis; detection;

```

```

cell typing; abnormal cell function; ss.
XX OS Homo sapiens.
XX PN WO9514772-A1.
XX PD 01-JUN-1995.
XX PF 11-NOV-1994; 94WO-JP01916.
XX PR 12-NOV-1993; 93JP-0355504.
XX PA (MATS/) MATSUBARA K.
XX PA (OKUB/) OKUBO K.
XX PI Matsubara K, Okubo K;
XX DR WPI; 1995-206931/27.
XX
XX Identifying gene signatures in 3'-directed human cDNA library - e.g.
XX for diagnosis of abnormal cell function, by preparing cDNA that
XX reflects relative abundance of corresp. mRNA in specific human
XX tissues
XX
XX Claim 1; Page 1582; 2245pp; Japanese.
XX
XX A single-stranded DNA (or its complementary strand or the corresp.
XX double-stranded DNA) which comprises one of the 7837 "Gs" sequences
XX given in AAT19001-r26837 and which is able to hybridise to part of
XX human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
XX sequences were obtained from 3'-directed cDNA libraries prepared
XX from various human tissues; synthesis of cDNA was initiated from the
XX 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
XX untranslated sequence is unique to a particular mRNA species, almost
XX all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
XX is constructed so as to reflect accurately the relative abundance of
XX different mRNAs in the particular tissue from which it was derived.
XX The appearance frequency of a given GS in a cDNA library can be
XX determined (esp. using primers and probes derived from the GS
XX sequences) as a means of diagnosing abnormal cell function or for
XX recognising different cell types.
XX
XX Sequence 126 BP; 44 A; 16 C; 25 G; 34 T; 7 other;

Alignment Scores:
Pred. No.: 7.9e-06 Length: 126
Score: 117.00 Matches: 20
Percent Similarity: 78.57% Conservative: 2
Best Local Similarity: 71.43% Mismatches: 6
Query Match: 22.41% Indels: 0
DB: 16 Gaps: 0

US-09-830-244B-1 (1-95) x AAT24320 (1-126)
QY 1 MetTrpPheHisLysValGlyArgLysGlnHisPheLysValThrPheTrpGluThrAsp 20
Db 26 ATGTGTTTCATCAAGTGGGAGAGAAAGACGACAAATTTAAAAATANCTTTTGGGNGACTGAN 85
QY 21 LeuSerAsnAsnLysThrLeuVal 28
Db 86 TTGAGTAATAATAAANCTTCAGTC 109
RESULT 8
ABL54353
ID ABL54353 standard; DNA; 9881 BP.
XX AC ABL54353;
XX DT 29-JUL-2002 (first entry)
XX DE Chemically treated apoptosis gene #27.
XX KW Apoptosis; HIV; Bloom syndrome; cardiopathy;

```

```

XX OS Homo sapiens.
XX PN WO9845436-A2.
XX PD 15-OCT-1998.
XX PF 10-APR-1998; 98WO-US06955.
XX PR 10-APR-1997; 97US-0838821.
XX PA (GEM) GENETICS INST INC.
XX PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
XX PI Racie LA, Spaulding V, Treacy M;
XX DR WPI; 1999-070077/06.
XX PT New polynucleotides encoding human secreted proteins - derived from
XX PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
XX PT ovary, pituitary, retina and colon cDNA libraries.
XX PS Claim 1; Page 245; 618pp; English.
XX CC The present sequence represents a human expressed sequence tag (EST).
XX CC The polynucleotide, which is a secreted EST, and the encoded protein
XX CC are predicted to have useful biological activities which would make
XX CC them suitable for treating, preventing or ameliorating medical
XX CC conditions in humans and animals, although no supporting data is
XX CC given. Suggested activities include nutritional activity, immune
XX CC stimulating or suppressing activity, haematopoiesis regulating
XX CC activity, tissue growth activity, activin/inhibin activity,
XX CC chemotactic/chemokinetic activity, activin/inhibin activity,
XX CC activity, receptor/ligand activity, haemostatic and thrombolytic
XX CC adherin/tumour invasion suppressor activity, tumour inhibition
XX CC activity. The polynucleotide may also be useful for gene therapy.
XX SQ Sequence 516 BP; 156 A; 181 C; 97 G; 82 T; 0 other;

Alignment Scores:
Pred. No.: 1 33e-05 Length: 516
Score: 121.50 Matches: 28
Percent Similarity: 61.02% Conservative: 8
Best Local Similarity: 47.46% Mismatches: 16
Query Match: 23.28% Indels: 7
DB: 2 Gaps: 2

US-09-830-244B-1 (1-95) x AAV89525 (1-516)
QY 30 LeuLysLysLysProPheHisLeuTyrCysValIleTyrIlePro----- 45
Db 41 TTGAGTAAAGAGAACCCAGCCCAATATGAGTTCTCTTTCCTTCATCTTGGCTC 100
QY 46 --LeuValProLysLeuIleLeuPheLeuAspIleAlaPheIleProLysSerLeu 64
Db 101 TCATGGTTTCATGATGGAGCTGATCATCTG-----AAGAGTATGGGTATGGCCTT 154
QY 65 IleserGlnPheGlnAsnHisTyrThrHisAsnHisThrAsnHisAsnThrAsn 83
Db 155 ATCAGCAGTTCAGACACACCACTATACCCACACCACTACCAACCAACCAACCAAC 211

RESULT 7
AAT24320
ID AAT24320 standard; cDNA to mRNA; 126 BP.
XX AC AAT24320;
XX AC AAT24320;
XX DT 22-SEP-1996 (first entry)
XX DE Human gene signature HUMGS06347.
XX KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
human; cloning; mapping; non-biased library; diagnosis; detection;

cell typing; abnormal cell function; ss.
Homo sapiens.
WO9514772-A1.
01-JUN-1995.
11-NOV-1994; 94WO-JP01916.
12-NOV-1993; 93JP-0355504.
(MATS/) MATSUBARA K.
(OKUB/) OKUBO K.
Matsubara K, Okubo K;
WPI; 1995-206931/27.
Identifying gene signatures in 3'-directed human cDNA library - e.g.
for diagnosis of abnormal cell function, by preparing cDNA that
reflects relative abundance of corresp. mRNA in specific human
tissues
Claim 1; Page 1582; 2245pp; Japanese.
A single-stranded DNA (or its complementary strand or the corresp.
double-stranded DNA) which comprises one of the 7837 "GS" sequences
given in AAT19001-T26837 and which is able to hybridise to part of
human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
sequences were obtained from 3'-directed cDNA libraries prepared
from various human tissues; synthesis of cDNA was initiated from the
3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
untranslated sequence is unique to a particular mRNA species, almost
all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
is constructed so as to reflect accurately the relative abundance of
different mRNAs in the particular tissue from which it was derived.
The appearance frequency of a given GS in a cDNA library can be
determined (esp. using primers and probes derived from the GS
sequences) as a means of diagnosing abnormal cell function or for
recognising different cell types.
SQ Sequence 126 BP; 44 A; 16 C; 25 G; 34 T; 7 other;

Alignment Scores:
Pred. No.: 7.9e-06 Length: 126
Score: 117.00 Matches: 20
Percent Similarity: 78.57% Conservative: 2
Best Local Similarity: 71.43% Mismatches: 6
Query Match: 22.41% Indels: 0
DB: 16 Gaps: 0

US-09-830-244B-1 (1-95) x AAT24320 (1-126)
QY 1 MetTrpPheHisLysValGlyArgLysGlnHisPheLysValThrPheTrpGluThrAsp 20
Db 26 ATGTGGTTTCATCAAGTGGGAAGAAAGACGAATTTAAATAATCTTTTGGGNGACTGAN 85
QY 21 LeuSerAsnLysThrLeuVal 28
Db 86 TTGAGTAATAAATCACTTCAGTC 109

RESULT 8
ABL54353
ID ABL54353 standard; DNA; 9881 BP.
XX AC ABL54353;
XX AC ABL54353;
XX DT 29-JUL-2002 (first entry)
XX DE Chemically treated apoptosis gene #27.
XX KW Apoptosis; HIV; Bloom syndrome; cardiopathy;

```









PR 07-APR-2000; 2000DE-1019173.  
 PR 30-JUN-2000; 2000DE-1032529.  
 PR 01-SEP-2000; 2000DE-1043826.  
 XX  
 PA (EPIG-) EPIGENOMICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K;  
 XX  
 DR WPI; 2001-602752/68.  
 XX  
 PT Fragments of chemically modified genes associated with tumour suppressor  
 PT genes and oncogenes, useful in designing primers and probes for  
 PT analysing diseases associated with cytosine methylation state e.g.  
 PT cancer  
 XX  
 PS Claim 1; SEQ ID No 48; 27pp; English.  
 XX  
 CC The invention relates to a nucleic acid comprising a sequence of 18  
 CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with  
 CC bisulphite, of genes associated with tumour suppression and  
 CC oncogenes having a sequence taken from 536 (actually 533 since  
 CC numbers 408, 438 and 500 are missing from the sequence listing) sequences  
 CC (SS) and sequences complementary to (SS). The nucleic acid may be a  
 CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may  
 CC form part of a set of probes for detecting the cytosine methylation state  
 CC and/or single nucleotide polymorphisms and also to be used in an  
 CC array for analysing diseases associated with CpG dinucleotides e.g.  
 CC cancers and tumours. The probes can also be used in a method for  
 CC ascertaining genetic and/or epigenetic parameters for the diagnosis  
 CC of diseases, by analysing cytosine methylation. The parameters may be  
 CC compared to another set of genetic and/or epigenetic parameters, the  
 CC differences serving as basis for diagnosis and/or prognosis events which  
 CC are disadvantageous to patients. The present sequence is one of the  
 CC 533 genomic sequences derived from tumour suppressor genes and  
 CC oncogenes. Sequences with even numbered Seq ID numbers are the  
 CC complementary sequences of the corresponding odd numbered sequence (e.g.  
 CC ID 2 and ID1, ID 536 and ID 535, except for those whose partner sequence  
 CC is missing).  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 9021 BP; 2455 A; 110 C; 2022 G; 4434 T; 0 other;

Alignment Scores:  
 Pred. No.: Length: 9021  
 Score: 52.00 Matches: 34  
 Percent Similarity: 52.38% Conservative: 10  
 Best Local Similarity: 40.48% Mismatches: 33  
 Query Match: 17.62% Indels: 7  
 DB: 22 Gaps: 3

US-09-830-244B-1 (1-95) x AAS45296 (1-9021)

QY 12 PheLysValThrPheTrpGluThrAspLeuSerAsn---AsnLysThrLeuValSerLeu 30  
 Db 6174 TTCAAAAATACACACTAAACACACTTATATATATATTAACAACTACTCTCAAAATAAT 6115  
 QY 31 LysLysLysLysProPheHisLeuTyrcysValleTyrlleProLeu-----ValPro 48  
 Db 6114 TCACAAAAAACAAT 6055  
 QY 49 LysLeuLeuLeuLeuPheLeuAspIleAla-PheIleProLysSerLeuIleSerGlnPh 68  
 Db 6054 CATTATCCATTTTAAACCTAAATTAACCTATTCATCCCA-----ATAACACAAAA 6004  
 QY 68 eGlnAsnHisTyrrHisAsnHsThrAsnHsAsnThrAsnAsnIleargPheAs 88  
 Db 6003 ATCTAATTCGCACATATCTTCACAAACATACAACTACAAATTAATTTTTTTCACAAAAA 5944  
 QY 88 nileIleSer 91

Db 5943 TATAATAACT 5934  
 RESULT 14  
 AAS45496/c

ID AAS45496 standard; DNA; 7038 BP.  
 XX  
 AC AAS45496;  
 XX

DT 18-DEC-2001 (first entry)  
 XX

DE Chemically pretreated genomic DNA associated with cell cycle #101.  
 XX

KW Cell cycle; human; CpG dinucleotide; cytosine methylation; HIV; aging;  
 KW human immunodeficiency virus; neurodegenerative disorder; solid tumour;  
 KW graft-versus-host disease; glomerular disease; Lewy body disease; cancer;  
 KW arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritic;  
 KW immunosuppressive; antitumour; cytostatic; antiarteriosclerotic; ds;  
 KW PCR primer.  
 XX

OS Homo sapiens.  
 XX

PN WO200168911-A2.  
 XX

PD 20-SEP-2001.  
 XX

PF 15-MAR-2001; 2001WO-EP02945.  
 XX

PR 15-MAR-2000; 2000DE-1013847.  
 PR 06-APR-2000; 2000DE-1019058.  
 PR 07-APR-2000; 2000DE-1019173.  
 PR 30-JUN-2000; 2000DE-1032529.  
 PR 01-SEP-2000; 2000DE-1043826.  
 XX

PA (EPIG-) EPIGENOMICS AG.  
 XX

PI Olek A, Piepenbrock C, Berlin K;  
 XX

DR WPI; 2001-602751/68.  
 XX

PT Designing primers and probes for analysing diseases associated with  
 PT cytosine methylation state e.g. arthritis, cancer, aging  
 PT arteriosclerosis comprising fragments of chemically modified genes  
 PT associated with cell cycle -  
 XX

PS Claim 1; SEQ ID No 201; 28pp; English.  
 XX

CC Sequences AAS45296-AAS45520 represent chemically pretreated genomic DNA  
 CC molecules associated with the cell cycle and specific PCR primers of the  
 CC invention. The sequences are useful for detecting the methylation state  
 CC of all CpG dinucleotides in a sequence and therefore for analysing  
 CC associated diseases. By analysing cytosine methylation in the pretreated  
 CC DNA, genetic and/or epigenetic parameters for the diagnosis and therapy  
 CC of existing diseases or the predisposition to specific diseases can be  
 CC ascertained. The parameters may be compared to another set of genetic  
 CC and/or epigenetic parameters, the differences serving as basis for  
 CC diagnosis and/or prognosis events which are disadvantageous to patients.  
 CC The sequences of the invention are useful for the diagnosis and therapy  
 CC of HIV infection, neurodegenerative disorders, graft-versus-host disease,  
 CC aging, glomerular disease, Lewy body disease, arthritis,  
 CC arteriosclerosis, solid tumours and cancers.  
 XX

SQ Sequence 7038 BP; 1424 A; 375 C; 2062 G; 3177 T; 0 other;

Alignment Scores:

Pred. No.: 1.52 Length: 7038  
 Score: 91.50 Matches: 37  
 Percent Similarity: 38.89% Conservative: 12  
 Best Local Similarity: 29.37% Mismatches: 35  
 Query Match: 17.53% Indels: 42  
 DB: 22 Gaps: 5

US-09-830-244B-1 (1-95) x AAS45496 (1-7038)



QY 3 PheHisLysValGlyArgLysGlnHisPheLys-ValThrPheTrpGluThrAspLeuSe 22  
 Db 4931 TTTAAAACTAAACAAAAAACCACTTAAACCAACGTC---GAAAAAACCTTAA 4875  
 QY 22 rAsnAsn-----LysThrLeuValSerLeuLysLysLysPr 35  
 Db 4874 CAACAAATAAACATATCCATCTCTACAAACATTTAAATAATATAAAAAA 4815  
 QY 35 O-----PheHisLeuTyrCy 40  
 Db 4814 AAAAAACCTACCTCTACAAATATCAATAAATAAATAATATACATACAAAA 4755  
 QY 40 sValIleTyr-----IleProLeuValProLysLeuLeileLeuPheLeuAs 56  
 Db 4754 CTTAAATATATTTCAAAAAAATAAATAATTTAACTTAAACCAACCACTTAA 4875  
 QY 56 pIleAlaPheIleProLysSerLeuIleSerGlnPheGlnAsn----- 70  
 Db 4694 ATTAACCTCAATTCCTCCCATATCTCTACATCTCTTCAAAATCCAAAAA 4635  
 QY 71 -----AsnHisTyrThrHisAsnHisThrAsnHisAsnth 82  
 Db 4634 ACGAATACTTTAATTCAAAAAATAAATAATTTAACTTAAACCAACCACTTAA 4875  
 QY 82 rAsnAsnIleArgPhe 87  
 Db 4574 GAAAAACCTCTCTCTTC 4559

RESULT 15  
 ABL34097/c  
 ID ABL34097 standard; DNA; 7038 BP.  
 XX AC ABL34097;  
 XX DT 26-MAR-2002 (first entry)  
 DE Human immune system associated gene SEQ ID NO: 2070.  
 KW Human; immune system disease; cytosine methylation; antiasthmatic;  
 KW antiarteriosclerotic; anti-anemic; cytosine; cytosine; cytosine;  
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
 KW gene; ds.  
 OS Homo sapiens.  
 XX PN W0200200928-A2.  
 XX PD 03-JAN-2002.  
 XX PF 02-JUL-2001; 2001WO-EP07537.  
 XX PR 30-JUN-2000; 2000DE-1032529.  
 XX PR 01-SEP-2000; 2000DE-1043826.  
 XX PA (EPIG-) EPIGENOMICS AG.  
 XX PI Olek A, Piepenbrock C, Berlin K;  
 XX WI: 2002-130909/17.  
 XX Nucleic acid comprising fragment of chemically modified gene, useful  
 XX for diagnosis and treatment of diseases associated with abnormal  
 XX cytosine methylation -  
 XX ClaIm 1; SEQ ID NO 2070; 32pp + Sequence Listing; German.  
 XX The present invention provides a number of human immune system associated  
 XX genes which are modified by the methylation of cytosines. The sequences

CC can be used in the diagnosis and treatment of immune system disorders,  
 CC including eye diseases such as retinopathy, neovascular glaucoma and  
 CC macular degeneration, arteriosclerosis, anemia, cancer, acute myeloid  
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
 CC diseases. The present sequence is a gene of the invention.  
 XX

SQ Sequence 7038 BP; 1424 A; 375 C; 2062 G; 3177 T; 0 other;

## Alignment Scores:

Pred. No.: 1.52 Length: 7038  
 Score: 91.50 Matches: 37  
 Percent Similarity: 38.89% Conservative: 12  
 Best Local Similarity: 29.37% Mismatches: 35  
 Query Match: 17.53% Indels: 42  
 DB: 24 Gaps: 5

US-09-830-244B-1 (1-95) x ABL34097 (1-7038)

QY 3 PheHisLysValGlyArgLysGlnHisPheLys-ValThrPheTrpGluThrAspLeuSe 22  
 Db 4931 TTTAAAACTAAACAAAAAACCACTTAAACCAACGTC---GAAAAAACCTTAA 4875  
 QY 22 rAsnAsn-----LysThrLeuValSerLeuLysLysLysPr 35  
 Db 4874 CAACAAATAAACATATCCATCTCTACAAACATTTAAATAATATAAAAAA 4815  
 QY 35 O-----PheHisLeuTyrCy 40  
 Db 4814 AAAAAACCTACCTCTACAAATATCAATAAATAAATAATATACATACAAAA 4755  
 QY 40 sValIleTyr-----IleProLeuValProLysLeuLeileLeuPheLeuAs 56  
 Db 4754 CTTAAATATATTTCAAAAAAATAAATAATTTAACTTAAACCAACCACTTAA 4695  
 QY 56 pIleAlaPheIleProLysSerLeuIleSerGlnPheGlnAsn----- 70  
 Db 4694 ATTAACCTCAATTCCTCCCATATCTCTACATCTCTTCAAAATCCAAAAA 4635  
 QY 71 -----AsnHisTyrThrHisAsnHisThrAsnHisAsnth 82  
 Db 4634 ACGAATACTTTAATTCAAAAAATAAATAATTTAACTTAAACCAACCACTTAA 4575  
 QY 82 rAsnAsnIleArgPhe 87  
 Db 4574 GAAAAACCTCTCTCTTC 4559

Search completed: April 14, 2003, 23:13:50  
 Job time : 232 secs





1



```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/385.335A
; FILING DATE: 08-FEB-1995
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: WHI95-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6350 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
US-08-385-335A-9
Alignment Scores:
Pred. No.: 41.9 Length: 6350
Score: 72.00 Matches: 14
Percent Similarity: 54.90% Conservative: 14
Best Local Similarity: 27.45% Mismatches: 23
Query Match: 13.79% Indels: 0
DB: 2 Gaps: 0

US-09-830-244B-1 (1-95) x US-08-385-335A-9 (1-6350)
QY 3 PheHisLysValGlyArgLysGlnHisPheLysValThrPheTpgLutThrAspLeuSer 22
Db 2397 TTTTCTAAATAAATACAGAGCTTTTCCCTTACTTTCTGGGGGACATCCAAAC 2456
QY 23 AsnAsnLysThrLeuValSerLeuLysLysLysProPheHisLeuTyrCysValle 42
Db 2457 CGTTCGATCGAATTCATTCGGGTAAACGGTTCCTGTTTATTTATGACTCTTATC 2516
QY 43 TyrileProLeuValProLysLeuIleLeu 53
Db 2517 TCCTCAAGTCAGTTCAGTTCCTGCTACCTT 2549

RESULT 8
US-09-521-526-2/c
; Sequence 2, Application US/09521526
; Patent No. 6290965
; GENERAL INFORMATION:
; APPLICANT: JANSEN, KATHRIN U.
; APPLICANT: HOFMANN, KATHRYN J.
; TITLE OF INVENTION: DNA ENCODING HUMAN PAPILLOMA VIRUS TYPE
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CHRISTINE E. CARTY
; STREET: 126 E. LINCOLN AVENUE - P.O. BOX 2000
; CITY: RAHWAY
; STATE: NJ
; COUNTRY: USA
; ZIP: 07065-0900
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/521,526
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/310,468
; FILING DATE: 22-SEP-1994

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/385.335A
; FILING DATE: 08-FEB-1995
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: CARTY, CHRISTINE E.
; REGISTRATION NUMBER: 36,099
; REFERENCE/DOCKET NUMBER: 19307
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908) 594-6734
; TELEFAX: (908) 594-4720
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8010 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-09-521-526-2
Alignment Scores:
Pred. No.: 88.1 Length: 8010
Score: 70.50 Matches: 25
Percent Similarity: 43.88% Conservative: 18
Best Local Similarity: 25.51% Mismatches: 31
Query Match: 13.51% Indels: 24
DB: 4 Gaps: 5

US-09-830-244B-1 (1-95) x US-09-521-526-2 (1-8010)
QY 3 PheHisLysValGlyArg-----LysGlnHisPheLysValThrPheTpgLutThrAsp 20
Db 7717 TGGCACAATAATATAGGATTAACAGGCAACACGCTTCTTAACAAATTCGTGGAAGTGTAT 7658
QY 21 LeuSerAsnAsnLysThrLeuValSerLeuLysLysLysProPheHisLeuTyrCys 40
Db 7657 GCCAAGGCGCAACCGAAACCGTTGCAATATATAAAA----- 7622
QY 41 ValletYrileProLeuValProLysLeuIleLeuPheLeuAspIleAlaPheile 60
Db 7621 -----TATATT---ATTGTAACAAAAATATATATATGTCATGTATATTACTTATGTG 7571
QY 61 ProLysSerLeuIleSerGlnPheGln---AsnAsnHis----- 72
Db 7570 TGTAAAGCGCGTGCACAAACAAACGCGTGCACACGCCACTGAGTCACAGGGTGTAGTTAA 7511
QY 73 -----TyrThrHisAsnHisThr-AsnHisAsnThrAsnAsnIleArgPhe 87
Db 7510 TTCTTTTTCACATTAACACACATACATACATACACAAATAAATACATTAC 7459

RESULT 9
PCT-US95-11859-2/c
; Sequence 2, Application PC/TUS9511859
; GENERAL INFORMATION:
; APPLICANT: JANSEN, KATHRIN U.
; APPLICANT: HOFMANN, KATHRYN J.
; TITLE OF INVENTION: DNA ENCODING HUMAN PAPILLOMA VIRUS TYPE
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CHRISTINE E. CARTY
; STREET: 126 E. LINCOLN AVENUE - P.O. BOX 2000
; CITY: RAHWAY
; STATE: NJ
; COUNTRY: USA
; ZIP: 07065-0907
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/11859
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/310,468
```

; FILING DATE: 22-SEP-1994  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: CARTY, CHRISTINE E.  
 ; REGISTRATION NUMBER: 36,099  
 ; REFERENCE/DOCKET NUMBER: 19307 PCT  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (908) 594-6734  
 ; TELEFAX: (908) 594-4720  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 8010 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: cDNA  
 PCT-US95-11859-2

Alignment Scores:  
 Pred. No.: 88.1 Length: 8010  
 Score: 70.50 Matches: 25  
 Percent Similarity: 43.88% Conservative: 18  
 Best Local Similarity: 25.51% Mismatches: 31  
 Query Match: 13.51% Indels: 24  
 DB: 5 Gaps: 5

US-09-830-244B-1 (1-95) x PCT-US95-11859-2 (1-8010)

QY 3 PheHisLysValGlyArg-----LysGlnHisPheLysValThrPheTrpGluThrAsp 20  
 DB 7717 TGGCACAAATATATAGATTAAACAGGCAACACGTTGTGAACAAATGGTGGAAAGTGAT 7658  
 QY 21 LeuSerAsnAsnLysThrLeuValSerLeuLysLysLysLysPropHeHisLeuTyrCys 40  
 DB 7657 GCCAGGCGCACCGCAACCGGTGGCAATATATAAAA----- 7622  
 QY 41 ValIleTyrIleProLeuValProLysLeuIleLeuPheLeuAspIleAlaPheIle 60  
 DB 7621 -----TATATT---ATTGTAACAAATATATATATTGTGCATGTATATTACTTATGTG 7571  
 QY 61 ProLysSerLeuIleSerGlnPheGln---AsnAsnHis----- 72  
 DB 7570 TGTAGGCGCGTGCACAAACGCGTGCACACGCCACTGAGTCACAGGCGTAGTTAA 7511  
 QY 73 -----TyrThrHisAsnHisThr-AsnHisAsnThrAsnAsnIleArgPhe 87  
 DB 7510 TTGTTATTGCACATTAACACACATACATACATACAAATAAATTACATTAC 7459

RESULT 10

US-08-998-416-915  
 ; Sequence 915, Application US/08998416  
 ; Patent No. 6239264

; GENERAL INFORMATION:  
 ; APPLICANT: Philippsen, Peter  
 ; APPLICANT: Pohlmann, Rainer  
 ; APPLICANT: Steiner, Sabine  
 ; APPLICANT: Mohr, Christine  
 ; APPLICANT: Wendland, Jürgen  
 ; APPLICANT: Knechtle, Philipp  
 ; APPLICANT: Reibschung, Corinne  
 ; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPHII  
 ; NUMBER OF INVENTION: AND USES THEREOF  
 ; NUMBER OF SEQUENCES: 1152  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: No. 6239264artis Corporation  
 ; STREET: 3054 Cornwallis Road  
 ; CITY: Research Triangle Park  
 ; STATE: No. 6239264th Carolina  
 ; COUNTRY: USA  
 ; ZIP: 27709

; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/998,416  
 ; FILING DATE: 24-DEC-1997  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: CH 0016/97  
 ; FILING DATE: 31-DEC-1996  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Meigs, J. Timothy  
 ; REGISTRATION NUMBER: 38,241  
 ; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 919-541-8587  
 ; TELEFAX: 919-541-8689

INFORMATION FOR SEQ ID NO: 915:

; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 688 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: PAG1570RP

US-08-998-416-915

Alignment Scores:  
 Pred. No.: 5.33 Length: 688  
 Score: 69.00 Matches: 24  
 Percent Similarity: 42.68% Conservative: 11  
 Best Local Similarity: 29.27% Mismatches: 41  
 Query Match: 13.22% Indels: 6  
 DB: 4 Gaps: 2

US-09-830-244B-1 (1-95) x US-08-998-416-915 (1-688)

QY 9 LysGlnHisPheLysValThrPheTrpGluThrAspLeuSerAsnAsnLysThrLeuVal 28  
 DB 38 AAAGACAGATTCCACAGTCCACACACTTGCACAAACCATCTGTAAGCTTTCAACGATTGCGC 97  
 QY 29 SerLeuLysLysLysLysLysProPheHisLeuTyrCysValIleTyrIleProLeuValPro 48  
 DB 98 AGCAGCAACACCCAGCAACCGCTCTCTCAATCTCATCTCCACGACCAACCAAGGTTCCAC 157  
 QY 49 LysLeuIleLeuLeuPheLeuAspIleAlaPheIleProLysSerLeuIleSerGlnPhe 68  
 DB 158 AAGCTCAGCAGCAGCTACTTCAACAACAAGGTTCCACAGCTCAGCAGCAGCTACTTTC 217  
 QY 69 GlnAsnAsn-----HisTyrThrHisAsnHisThrAsn---HisAsnThr 82  
 DB 218 AACAAACAACAACAACAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 277  
 QY 83 AsnAsn 84  
 DB 278 CACAAC 283

RESULT 11

US-09-272-114-2/c  
 ; Sequence 2, Application US/09272114  
 ; Patent No. 6133013

; GENERAL INFORMATION:  
 ; APPLICANT: Viola, Ronald E.  
 ; APPLICANT: Jayasekera, Maithri M. K.  
 ; APPLICANT: Saribas, Abdullah S.  
 ; TITLE OF INVENTION: Truncated Aspartase Enzyme Derivatives  
 ; TITLE OF INVENTION: and Uses Thereof  
 ; NUMBER OF SEQUENCES: 2  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Renner, Kenner, Greive, Bobak,  
 ; ADDRESSEE: Taylor & Weber  
 ; STREET: 1610 First National Tower  
 ; CITY: Akron  
 ; STATE: Ohio







```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/869,933
; FILING DATE: 19920416
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 40399/154 NIHD
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11298 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; ORIGINAL SOURCE:
; ORGANISM: homo sapien
; STRAIN: FCRI beta
; US-07-869-933-31

```

```

Alignment Scores:
Pred. No.: 215
Score: 69.00
Percent Similarity: 40.00%
Best Local Similarity: 31.76%
Query Match: 13.22%
DB: 1
Gaps: 3

```

US-09-830-244B-1 (1-95) x US-07-869-933-31 (1-11298)

```

QY 7 GlyArgLysGlnHisPheLysValThrPheTrpGluThrAspLeuSerAsnLys--- 25
Db 5669 GGTGAGAAACAGCATCATCTACTCAATTTCTTA-----TTGACACGACACAAATGA 5619
QY 26 -----ThrLeuValSerLeuLysLysLysLysPropheHisLeuTyrCys 40
Db 5618 CAACAAAAAGATCATTTGGTA----- 5595
QY 41 ValIleTyrIleProLeuValProLysLeuIleLeuPheLeuAspIleAlaPheIle 60
Db 5594 -----ITCGCCCCAAAGATTATTTCTTATAGCATCTCTCTTTTCATC 5550
QY 61 -ProLysSerLeuIleSerGlnPheGlnAsnHis-TyrThrHisAsnHisThrAsnH 80
Db 5549 TCCTTACCCTGACGACGCTTTTAAAGTGACACATACACACACACACACACACA 5490
QY 80 isAsnThrAsn 83
Db 5489 CACACACAAAC 5479

```

RESULT 14

```

US-08-201-879A-2/c
; Sequence 2, Application US/08201879A
; Patent No. 5807988
; GENERAL INFORMATION:
; APPLICANT: KINET, Jean-Pierre
; APPLICANT: JOUVIN, Marie-Helene
; TITLE OF INVENTION: ISOLATION, CHARACTERIZATION, AND USE OF
; TITLE OF INVENTION: THE HUMAN B SUBUNIT OF THE HIGH AFFINITY RECEPTOR FOR
; TITLE OF INVENTION: IMMUNOGLOBULIN E
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/201,879A
; FILING DATE: 24-FEB-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/869,933
; FILING DATE: 16-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/03419
; FILING DATE: 16-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 40399/234/NIHD
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11298 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; STRAIN: FCRI beta
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(456..511, 1381..1510, 2026..2160, 4475..4531,
; LOCATION: 5079..5237, 5640..5738, 7224..7319)
; US-08-201-879A-2

```

```

Alignment Scores:
Pred. No.: 215
Score: 69.00
Percent Similarity: 40.00%
Best Local Similarity: 31.76%
Query Match: 13.22%
DB: 1
Gaps: 3

```

US-09-830-244B-1 (1-95) x US-08-201-879A-2 (1-11298)

```

QY 7 GlyArgLysGlnHisPheLysValThrPheTrpGluThrAspLeuSerAsnLys--- 25
Db 5669 GGTGAGAAACAGCATCATCTACTCAATTTCTTA-----TTGACACGACACAAATGA 5619
QY 26 -----ThrLeuValSerLeuLysLysLysLysPropheHisLeuTyrCys 40
Db 5618 CAACAAAAAGATCATTTGGTA----- 5595
QY 41 ValIleTyrIleProLeuValProLysLeuIleLeuPheLeuAspIleAlaPheIle 60
Db 5594 -----ITCGCCCCAAAGATTATTTCTTATAGCATCTCTCTTTTCATC 5550
QY 61 -ProLysSerLeuIleSerGlnPheGlnAsnHis-TyrThrHisAsnHisThrAsnH 80
Db 5549 TCCTTACCCTGACGACGCTTTTAAAGTGACACATACACACACACACACACA 5490
QY 80 isAsnThrAsn 83
Db 5489 CACACACAAAC 5479

```

RESULT 15

```

US-09-103-663-31/c
; Sequence 31, Application US/09103663D
; Patent No. 6171803
; GENERAL INFORMATION:

```

APPLICANT: Kinet et al.  
TITLE OF INVENTION: Isolation, characterization, and use of the human beta  
TITLE OF INVENTION: subunit of the high affinity receptor for  
TITLE OF INVENTION: immunoglobulin E.  
FILE REFERENCE: 50490  
CURRENT APPLICATION NUMBER: US/09/103,663D  
CURRENT FILING DATE: 1998-06-23  
EARLIER APPLICATION NUMBER: 07/869,933  
EARLIER FILING DATE: 1992-04-16  
NUMBER OF SEQ ID NOS: 35  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 31  
LENGTH: 11298  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-103-663-31

Alignment Scores:  
Pred. NO.: 215 Length: 11298  
Score: 69.00 Matches: 27  
Percent Similarity: 40.00% Conservative: 7  
Best Local Similarity: 31.76% Mismatches: 23  
Query Match: 13.22% Indels: 28  
DB: 4 Gaps: 3

US-09-830-244B-1 (1-95) x US-09-103-663-31 (1-11298)

Qy 7 GlyArgLysGlnHisPheLysValThrPheTrpGluThrAspLeuSerAsnAsnLys--- 25  
Db 5669 GGTGAGAAACAGCATCATCATCAATTCCTA-----TTGAACAGCAACAAATGA 5619  
Qy 26 -----ThrLeuValSerLeuLysLysLysLysProPheHisLeuTyrCys 40  
Db 5618 CAAACAAAAGAGTCACATTGGTA----- 5595  
Qy 41 ValIleTyrIleProLeuValProLysLeuIleIleLeuPheLeuAspIleAlaPheIle 60  
Db 5594 -----TTGCGCCCAAAAGTTTATTCTTATAGCATCTCCTGTTTCATC 5550  
Qy 61 -ProLysSerLeuIleSerGlnPheGlnAsnAsnHis-TyrThrHisAsnHisThrAsnH 80  
Db 5549 TCCCTACCATCTGACCATGCTCTTTAAAGTGACACATACACACACACACACACACA 5490  
Qy 80 IsAsnThrAsn 83  
Db 5489 CACACACAAAC 5479

Search completed: April 15, 2003, 00:03:56  
Job time : 104 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: April 14, 2003, 23:09:51 ; Search time 85 Seconds  
(without alignments)  
980.364 Million cell updates/sec

Title: US-09-830-244B-1

Perfect score: 522

Sequence: 1 MWFKVGRKQHKFTWEDT.....NHTNNTNIRFNISNCR 95

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 593429 seqs, 439583890 residues

Total number of hits satisfying chosen parameters: 1186858

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp  
-Q=/cgn2\_1/USPTO\_spool/US09830244/runat\_10042003\_090017\_27901/app\_query.fasta\_1.263  
-DB=PublishedApplications\_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1  
-LOOCL=0 -LOOEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62  
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEA=SIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US09830244 -CGEN\_1\_1.80 -runat\_10042003\_090017\_27901  
-NCPU=6 -ICPU=3 -NO\_XLPXY -NO\_XMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

PublishedApplications\_NA:  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	83.5	16.0	12571	10	US-09-965-553-20
C 2	79	15.1	2000	9	US-09-938-842A-4834
C 3	79	15.1	56737	10	US-09-782-378A-17
C 4	77.5	14.8	5109	9	US-09-422-569-11
C 5					Sequence 20, Appl
C 6					Sequence 4834, Ap
C 7					Sequence 17, Appl
C 8					Sequence 11, Appl

5	75	14.4	1503841	9	US-09-946-807-1	Sequence 1, Appl1
6	75	14.4	1503841	10	US-09-795-668-1	Sequence 1, Appl1
7	75	14.4	1503841	10	US-09-795-668-1	Sequence 1, Appl1
C 8	74.5	14.3	32038	9	US-10-079-854-292	Sequence 292, App
C 9	74.5	14.3	32038	10	US-09-764-878-292	Sequence 292, App
10	73	14.0	1248	10	US-09-841-132-358	Sequence 358, App
11	73	14.0	1278	10	US-09-841-132-358	Sequence 358, App
12	73	14.0	1311	10	US-09-841-132-359	Sequence 359, App
13	72	13.8	6700	9	US-09-938-842A-4705	Sequence 4705, Ap
14	72	13.8	6700	9	US-09-938-842A-4705	Sequence 4705, Ap
15	72	13.8	8161	10	US-09-759-152-3	Sequence 3, Appl1
16	72	13.8	8175	10	US-09-759-152-9	Sequence 9, Appl1
17	72	13.8	8518	10	US-09-759-152-9	Sequence 9, Appl1
18	71	13.6	2000	9	US-09-938-842A-3444	Sequence 7, Appl1
19	70.5	13.5	151	10	US-09-864-761-22646	Sequence 3444, Ap
20	70.5	13.5	492	10	US-09-864-761-22646	Sequence 22646, A
21	70.5	13.5	99014	10	US-09-864-761-5885	Sequence 5885, Ap
22	70.5	13.5	99014	10	US-09-864-761-5885	Sequence 5885, Ap
23	70.5	13.5	465237	10	US-09-880-107-3428	Sequence 3428, Ap
C 24	70	13.4	239	10	US-09-933-267A-1	Sequence 1, Appl1
C 25	70	13.4	239	10	US-09-933-267A-1	Sequence 1, Appl1
C 26	70	13.4	2563	9	US-10-072-349-250	Sequence 14, Appl1
C 27	70	13.4	2563	9	US-10-072-349-250	Sequence 250, App
C 28	70	13.4	2563	9	US-10-072-349-251	Sequence 251, App
C 29	69.5	13.3	943	10	US-09-764-855-251	Sequence 251, App
C 30	69.5	13.3	943	10	US-09-764-855-251	Sequence 251, App
C 31	69.5	13.3	1200	10	US-09-070-927A-527	Sequence 527, App
C 32	69.5	13.3	1200	10	US-09-070-927A-527	Sequence 527, App
C 33	69	13.2	2886	9	US-09-887-576-722	Sequence 2072, Ap
C 34	68.5	13.1	12541	10	US-09-801-368-131	Sequence 722, App
C 35	68.5	13.1	435	10	US-09-801-368-131	Sequence 2853, Ap
C 36	68.5	13.1	458	10	US-09-864-761-16771	Sequence 131, App
C 37	68.5	13.1	1863	10	US-09-864-761-16771	Sequence 2184, Ap
C 38	68.5	13.1	1863	10	US-09-770-444-474	Sequence 16771, A
C 39	68.5	13.1	2307	9	US-09-765-873A-13	Sequence 474, App
C 40	68.5	13.1	2658	9	US-10-278-173-160	Sequence 13, Appl
C 41	68.5	13.1	4325	9	US-09-858-525A-5	Sequence 160, App
C 42	68.5	13.1	30310	10	US-09-858-525A-6	Sequence 5, Appl1
C 43	68.5	13.1	152331	9	US-09-800-631-96	Sequence 6, Appl1
C 44	68.5	13.1	176373	9	US-10-095-407-16	Sequence 96, Appl
C 45	67.5	12.9	599	9	US-10-095-407-17	Sequence 16, Appl
C 46	67.5	12.9	599	9	US-09-933-267A-1	Sequence 17, Appl
C 47	67.5	12.9	599	9	US-09-933-267A-1	Sequence 1, Appl1
C 48	67.5	12.9	599	9	US-09-917-800A-881	Sequence 881, App
C 49	67.5	12.9	599	9	US-09-917-800A-881	Sequence 8191, Ap

#### ALIGNMENTS

RESULT 1  
US-09-965-553-20/C  
; Sequence 20, Application US/09965553  
; Patent No. US20020112259A1  
; GENERAL INFORMATION:  
; APPLICANT: Wright, David A.  
; TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto  
; FILE REFERENCE: P-1065 ISURF Plant Retroelement  
; CURRENT APPLICATION NUMBER: US/09/965,553  
; CURRENT FILING DATE: 2001-09-27  
; PRIOR APPLICATION NUMBER: 09/322,478  
; PRIOR FILING DATE: 1999-05-28  
; PRIOR APPLICATION NUMBER: 60/087125  
; PRIOR FILING DATE: 1998-05-29  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 20  
; LENGTH: 12571  
; TYPE: DNA  
; ORGANISM: Glycine max  
US-09-965-553-20

Alignment Scores:  
Pred. No.: 10.9  
Score: 83.50  
Percent Similarity: 44.29%  
Best Local Similarity: 31.43%  
Length: 12571  
Matches: 22  
Conservative: 9  
Mismatch: 28



; NUMBER OF SEQ ID NOS: 11  
 ; SOFTWARE: PatentIn ver. 2.1  
 ; SEQ ID NO 11  
 ; LENGTH: 5109  
 ; TYPE: DNA  
 ; ORGANISM: Caenorhabditis elegans  
 US-09-422-569-11

## Alignment Scores:

Pred. No.: 16.8 Length: 5109  
 Score: 77.50 Matches: 29  
 Percent Similarity: 40.00% Conservative: 13  
 Best Local Similarity: 27.62% Mismatches: 24  
 Query Match: 14.85% Indels: 39  
 DB: 9 Gaps: 5

US-09-830-244B-1 (1-95) x US-09-422-569-11 (1-5109)

QY 11 HisPhelysValThrPheTrpGluThrAspLeuSerAsnAsnLys----- 25  
 Db 1530 CATCAAAATTTGGTGTCTTTTGCAGAAAGATGTGGTTCAAACAAAGGTGCAATCATTTGA 1589  
 QY 26 -----ThrLeuValSerLeuLys 31  
 Db 1590 CTCATGTGGCGGTGTGTTCATAGCGACAGTGTATCATCATCCTTGGTGTGCTGAAG 1649  
 QY 32 LysLysLysProPheHisLeuTyrcysValIleTyrlleProLeuValProLysLeuIle 51  
 Db 1650 AAGAAACAG-----TACACATCCATTCATCATGTGTGGTGGAG----- 1688  
 QY 52 IleLeuPheLeuAspIleAlaPheIleProLysSerLeu---IleSerGlnPheGlnAsn 70  
 Db 1689 -----GTTGCGCGCTGTCCACCCAGAGGAGCGCCACCTGTCCAAGATGCAGCAG 1739  
 QY 71 AsnHisTyrrHisAsnHisThrAsnHisAsnThrAsnAsnIleArgPheAsnIleIle 90  
 Db 1740 AAGCGGTAC-----GAAATCAACCTACATAATCTTTG 1772  
 QY 91 SerAsnCysArgThr 95  
 Db 1773 AGCAGATGCAGAACT 1787

## RESULT 5

US-09-946-807-1  
 ; Sequence 1, Application US/09946807  
 ; Patent No. US20020165144A1  
 ; GENERAL INFORMATION:

; APPLICANT: Stefansson, Hreinn  
 ; APPLICANT: Steinthorsdottir, Valgerdur  
 ; APPLICANT: Gulcher, Jeffrey R.  
 ; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE  
 ; FILE REFERENCE: 2345.2004-001  
 ; CURRENT APPLICATION NUMBER: US/09/946,807  
 ; PRIOR FILING DATE: 2001-09-05  
 ; PRIOR APPLICATION NUMBER: US/09/795,668  
 ; PRIOR FILING DATE: 2001-02-28  
 ; PRIOR APPLICATION NUMBER: US 09/515,716  
 ; PRIOR FILING DATE: 2000-02-28  
 ; NUMBER OF SEQ ID NOS: 1531  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 1

LENGTH: 1503841

; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (1)...(1531)  
 ; OTHER INFORMATION: r-g or a  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (1)...(1531)  
 ; OTHER INFORMATION: y-t/u or c  
 ; FEATURE:

; NAME/KEY: misc\_feature  
 ; LOCATION: (1)...(1531)  
 ; OTHER INFORMATION: m-a or c  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (1)...(1531)  
 ; OTHER INFORMATION: k-g or t/u  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (1)...(1531)  
 ; OTHER INFORMATION: s-g or c  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (1)...(1531)  
 ; OTHER INFORMATION: w-a or t/u  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (1)...(1531)  
 ; OTHER INFORMATION: b-g or c or t/u  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (1)...(1531)  
 ; OTHER INFORMATION: d-a or g or t/u  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (1)...(1531)  
 ; OTHER INFORMATION: h-a or c or t/u  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (1)...(1531)  
 ; OTHER INFORMATION: v-a or g or c  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (1)...(1531)  
 ; OTHER INFORMATION: n-a or g or c or t/u  
 ; OTHER INFORMATION: n-a or g or c or t/u  
 US-09-946-807-1

## Alignment Scores:

Pred. No.: 1.46e+05 Length: 1503841  
 Score: 75.00 Matches: 21  
 Percent Similarity: 51.02% Conservative: 4  
 Best Local Similarity: 42.86% Mismatches: 18  
 Query Match: 14.37% Indels: 6  
 DB: 9 Gaps: 3

US-09-830-244B-1 (1-95) x US-09-946-807-1 (1-1503841)

QY 43 TyrIleProLeuValProLysLeuIleIleLeu-----PheLeuAspIleAlaPhe 59  
 Db 239178 TACATCCCATTCACCCCTCCATTTTATCATCTCTCTCTCT-----TGTTTC 239231  
 QY 60 IleProLysSerLeuIleSerGlnPheGlnAsnAsnHisTyrrHisAsnHisThrAsn 79  
 Db 239232 TTCTTCAAGTCCTCTCTCTCTCTCTCTTACAAAACACACACACACACACACAC 239288  
 QY 80 HisAsnThrAsnAsnIleArgPheAsn 88  
 Db 239289 ACACACACACACACACGACGACAAAGACAAAC 239315

## RESULT 6

US-09-795-668-1  
 ; Sequence 1, Application US/09795668  
 ; Patent No. US20020045577A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Stefansson, Hreinn  
 ; APPLICANT: Steinthorsdottir, Valgerdur  
 ; APPLICANT: Gulcher, Jeffrey R.  
 ; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE  
 ; FILE REFERENCE: 2345.2004-001  
 ; CURRENT APPLICATION NUMBER: US/09/795,668  
 ; CURRENT FILING DATE: 2001-02-28  
 ; PRIOR APPLICATION NUMBER: US 09/515,716  
 ; PRIOR FILING DATE: 2000-02-28



**Alignment Scores:**

```

Db 346 GCCTCT-----TCTTCTCGGAATCCTTATTGAACCTCAAGAACC 387
Qy 34 LysProPheHisLeu-----TyrCysValIleTyrIleProLeuVal 47
Db 388 GAGCCTCAACTATTGGCGATATTAGCCTCTCGTTTTCGTGTGTCGTCAGAAAT 447
Qy 48 ProLysLeuIleLeuPheLeuAspIleAa---PheIleProLysSerLeuIleSer 66
Db 448 CCT-----TTTCTTCTCTGAGACCTTTTATGCCAAAACCTATGTGCAA 492
Qy 67 -----GlnPheGlnAsnAsnHisTyrThrHisAsnHisThrAsnHisAsnThr 82
Db 493 GGGCAAAACGCTGCTTCTACAAAACCTTTTACTATGACCATGCGCATGTGGAATTA 552
Qy 83 AsnAsnIleArgPheAsnIleIleSerAsnCysArgThr 95
Db 553 CATTCATAGACTTTCGCATCATTCCTCCACATTTACACA 591

RESULT 11
US-09-841-132-538
; Sequence 538, Application US/09841132
; Patent No. US20020061848A1
; GENERAL INFORMATION:
; APPLICANT: Bhatia, Ajav
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Probst, Peter
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.469C8
; CURRENT APPLICATION NUMBER: US/09/841.132
; CURRENT FILING DATE: 2001-04-23
; NUMBER OF SEQ ID NOS: 599
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 538
; LENGTH: 1278
; TYPE: DNA
; ORGANISM: C. Trachomatis D serovar
US-09-841-132-538

Alignment Scores:
Pred. No.: 8.17 Length: 1278
Score: 73.00 Matches: 30
Percent Similarity: 43.36% Conservative: 19
Best Local Similarity: 26.55% Mismatches: 34
Query Match: 13.98% Indels: 30
DB: 10 Gaps: 7

US-09-830-244B-1 (1-95) x US-09-841-132-538 (1-1278)
Qy 2 TrpPheHis-----LysValGlyArg-----LysGlnHisPheLys 13
Db 325 TGGGAACACACTAAACAGCTGGCGGATATCCCTACTTTTGAAGAAGCTATCTTTTGA 384
Qy 14 ValThrPheTrpGluThrAspLeuSerAsnAsnLysThrLeuValSerLeuLysLys 33
Db 385 GCCTCT-----TCTTCTCAGAATCCTTATTGAACCTCAAGAACC 426
Qy 34 LysProPheHisLeu-----TyrCysValIleTyrIleProLeuVal 47
Db 427 GAGCCTCAACTATTGGCGATATTAGCCTCTCCGTTTTCGTGTGTCGTCAGAAAT 486
Qy 48 ProLysLeuIleLeuPheLeuAspIleAa---PheIleProLysSerLeuIleSer 66
Db 487 CCT-----TTTCTTCTCTGAGACCTTTTATGCCAAAACCTATGTGCAA 531
Qy 67 -----GlnPheGlnAsnAsnHisTyrThrHisAsnHisThrAsnHisAsnThr 82
Db 532 GGGCAAAACGCTGCTTCTACAAAACCTTTTACTATGACCATGCGCATGTGGAATTA 591
Qy 83 AsnAsnIleArgPheAsnIleIleSerAsnCysArgThr 95
Db 592 CATTCATAGACTTTCGCATCATTCCTCCACATTTACACA 630

```

```

RESULT 12
US-09-841-132-359
; Sequence 359, Application US/09841132
; Patent No. US20020061848A1
; GENERAL INFORMATION:
; APPLICANT: Bhatia, Ajav
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Probst, Peter
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.469C8
; CURRENT APPLICATION NUMBER: US/09/841.132
; CURRENT FILING DATE: 2001-04-23
; NUMBER OF SEQ ID NOS: 599
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 359
; LENGTH: 1311
; TYPE: DNA
; ORGANISM: Chlamydia
US-09-841-132-359

Alignment Scores:
Pred. No.: 8.48 Length: 1311
Score: 73.00 Matches: 30
Percent Similarity: 43.36% Conservative: 19
Best Local Similarity: 26.55% Mismatches: 34
Query Match: 13.98% Indels: 30
DB: 10 Gaps: 7

US-09-830-244B-1 (1-95) x US-09-841-132-359 (1-1311)
Qy 2 TrpPheHis-----LysValGlyArg-----LysGlnHisPheLys 13
Db 349 TGGGAACACACTAAACAGCTGGCGGATATCCCTACTTTTGAAGAAGCTATCTTTTGA 408
Qy 14 ValThrPheTrpGluThrAspLeuSerAsnAsnLysThrLeuValSerLeuLysLys 33
Db 409 GCCTCT-----TCTTCTCAGAATCCTTATTGAACCTCAAGAACC 450
Qy 34 LysProPheHisLeu-----TyrCysValIleTyrIleProLeuVal 47
Db 451 GAGCCTCAACTATTGGCGATATTAGCCTCTCCGTTTTCGTGTGTCGTCAGAAAT 510
Qy 48 ProLysLeuIleLeuPheLeuAspIleAa---PheIleProLysSerLeuIleSer 66
Db 511 CCT-----TTTCTTCTCTGAGACCTTTTATGCCAAAACCTATGTGCAA 555
Qy 67 -----GlnPheGlnAsnAsnHisTyrThrHisAsnHisThrAsnHisAsnThr 82
Db 556 GGGCAAAACGCTGCTTCTACAAAACCTTTTACTATGACCATGCGCATGTGGAATTA 615
Qy 83 AsnAsnIleArgPheAsnIleIleSerAsnCysArgThr 95
Db 616 CATTCATAGACTTTCGCATCATTCCTCCACATTTACACA 654

RESULT 13
US-09-838-842A-4705
; Sequence 4705, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINI
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938.842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647

```



; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/300,111  
; PRIOR FILING DATE: 2001-06-22  
; NUMBER OF SEQ ID NOS: 5379  
; SEQ ID NO 4705  
; LENGTH: 2000  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
US-09-938-842A-4705

Alignment Scores:  
Pred. No.: 21.2 Length: 2000  
Score: 72.00 Matches: 28  
Percent Similarity: 51.09% Conservative: 19  
Best Local Similarity: 30.43% Mismatches: 27  
Query Match: 13.79% Indels: 18  
DB: 9 Gaps: 5

US-09-830-244B-1 (1-95) x US-09-938-842A-4705 (1-2000)

QY 11 HisPheLysValThrPheTrpGluThrAsp-----LeuSerAsnLysThrLeuVal 28  
Db 344 AACTTCAAAATAGTCTATTGTTGACAAATCGTTATCCCAATAACAAATTTAGTA 403  
QY 29 SerLeuLysLysLys---LysPro--PheHisLeuTyrcysValIleTyrIleProLeuV 47  
Db 404 ATCATACCAAAATGTAGAACCTTAATTCATCTGCACCTGC-----GGCATACCATTTA 457  
QY 47 alProLysLeuIleLeuPheLeuAspIleAlaPhe-----59  
Db 458 AACCCGAATGATTAACTTATTTCATCTGCGATTTTCTTAGTTGAAATATG 517  
QY 60 --IleProLysSerLeuIleSerGlnPheGlnAsn-----AsnHisTyrThrHisA 76  
Db 518 GTTTACCTAAACACCGGAAAGAAAGAAAGTATGATGATTTATGCAATTTATGCAACACA 577  
QY 76 snHisThrAsnHisAsnThrAsnAsnIleArg 86  
Db 578 ACTACACAGGAACACACAGTAGAGATTANA 609

RESULT 14

US-09-759-152-3  
; Sequence 3, Application US/09759152  
; Patent No. US20020028433A1  
; GENERAL INFORMATION:  
; APPLICANT: Palmer, Michelle A.J.  
; APPLICANT: Gee, Melissa  
; APPLICANT: Tillotson, Bonnie  
; TITLE OF INVENTION: Systems for Sensitive Detection of G-Protein Coupled  
; TITLE OF INVENTION: Receptor and Orphan Receptor Function Using Reporter  
; FILE REFERENCE: 4085-235-27 CIP  
; CURRENT APPLICATION NUMBER: US/09/759,152  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 09/654,499  
; PRIOR FILING DATE: 2000-09-01  
; PRIOR APPLICATION NUMBER: US 60/180,669  
; PRIOR FILING DATE: 2000-02-07  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 3  
; LENGTH: 6700  
; TYPE: DNA  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: pICAST ALC.  
US-09-759-152-3

Alignment Scores:  
Pred. No.: 127 Length: 6700  
Score: 72.00 Matches: 14  
Percent Similarity: 54.90% Conservative: 14

Best Local Similarity: 27.45% Mismatches: 23  
Query Match: 13.79% Indels: 0  
DB: 10 Gaps: 0

US-09-830-244B-1 (1-95) x US-09-759-152-3 (1-6700)

QY 3 PheHisLysValGlyArgLysGlnHisPheLysValThrPheTrpGluThrAspLeuSer 22  
Db 5869 TTTTCTAAATAAATCAGAGGTCTTTTCCCCCTTACTTTCTGGGGTGGACATCCAAAC 5928  
QY 23 AsnAsnLysThrLeuValSerLeuLysLysLysLysLysLysLysLysLysLysLysLys 42  
Db 5929 CGTTCGATCGAATTCATTCGGGTAAACAGTTCCTGACCTTTTATGATGATGACTCTTATC 5988  
QY 43 TyrIleProLeuValProLysLeuIleLeu 53  
Db 5989 TCTTCAAGTCTAGTTCCAGTCTCTGTCTACCTT 6021

RESULT 15

US-09-759-152-9  
; Sequence 9, Application US/09759152  
; Patent No. US20020028433A1  
; GENERAL INFORMATION:  
; APPLICANT: Palmer, Michelle A.J.  
; APPLICANT: Gee, Melissa  
; APPLICANT: Tillotson, Bonnie  
; APPLICANT: Chang, Xiao-Jia  
; TITLE OF INVENTION: Systems for Sensitive Detection of G-Protein Coupled  
; TITLE OF INVENTION: Receptor and Orphan Receptor Function Using Reporter  
; FILE REFERENCE: 4085-235-27 CIP  
; CURRENT APPLICATION NUMBER: US/09/759,152  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 09/654,499  
; PRIOR FILING DATE: 2000-09-01  
; PRIOR APPLICATION NUMBER: US 60/180,669  
; PRIOR FILING DATE: 2000-02-07  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 9  
; LENGTH: 8161  
; TYPE: DNA  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: pICAST OMN.  
US-09-759-152-9

Alignment Scores:  
Pred. No.: 170 Length: 8161  
Score: 72.00 Matches: 14  
Percent Similarity: 54.90% Conservative: 14  
Best Local Similarity: 27.45% Mismatches: 23  
Query Match: 13.79% Indels: 0  
DB: 10 Gaps: 0

US-09-830-244B-1 (1-95) x US-09-759-152-9 (1-8161)

QY 3 PheHisLysValGlyArgLysGlnHisPheLysValThrPheTrpGluThrAspLeuSer 22  
Db 5513 TTTTCTAAATAAATCAGAGGTCTTTTCCCCCTTACTTTCTGGGGTGGACATCCAAAC 5572  
QY 23 AsnAsnLysThrLeuValSerLeuLysLysLysLysLysLysLysLysLysLysLysLys 42  
Db 5573 CGTTCGATCGAATTCATTCGGGTAAACAGTTCCTGACCTTTTATGATGATGACTCTTATC 5632  
QY 43 TyrIleProLeuValProLysLeuIleLeu 53  
Db 5633 TCTTCAAGTCTAGTTCCAGTCTCTGTCTACCTT 5665

Search completed: April 15, 2003, 00:13:14  
Job time : 559 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2003, 22:59:56 ; Search time 83 Seconds

(without alignments)  
110.033 Million cell updates/sec

Title: US-09-830-244B-1

Perfect score: 522

Sequence: 1 MFHKVGRKHQKFKVFWETD.....NHTNHTNNIRFNISNCR 95

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_73:\*

1: PIR1:\*

2: PIR2:\*

3: PIR3:\*

4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	76.5	14.7	2708	2 T09079	probable chloroqui
2	76.5	14.7	2819	2 T09080	probable chloroqui
3	76	14.6	215	2 F84508	probable CCAAT-box
4	76	14.6	580	2 T18439	hypothetical prote
5	76	14.6	600	2 T18446	hypothetical prote
6	75	14.4	163	2 T15985	hypothetical prote
7	73	14.0	426	2 D71552	probable oligopept
8	71	13.6	332	2 T25023	hypothetical prote
9	71	13.6	1274	2 S59405	probable membrane
10	70.5	13.5	567	2 S58750	NADH2 dehydrogenas
11	69.5	13.3	446	2 H90094	hypothetical prote
12	69.5	13.3	1017	2 T18488	hypothetical prote
13	69	13.2	454	2 T32974	protein F37B4.8 (i
14	69	13.2	665	2 E88379	GAF domain protei
15	69	13.2	1245	2 D71613	hypothetical prote
16	68.5	13.1	232	2 E90134	hypothetical prote
17	68.5	13.1	499	1 S30007	probable UTP-gluc
18	68	13.0	204	2 T82468	hypothetical prote
19	68	13.0	346	1 E64668	aspartate-semialde
20	68	13.0	683	2 AC0741	hypothetical prote
21	68	13.0	907	2 AE1352	hypothetical prote
22	67.5	12.9	346	2 A70144	membrane-associate
23	67.5	12.9	482	2 F97239	hypothetical prote
24	67	12.8	309	2 T24804	hypothetical prote
25	67	12.8	765	2 S72278	hypothetical prote
26	67	12.8	950	2 S72284	ATP-dependent Clp
27	67	12.8	1028	2 A56038	DNA-directed RNA p
28	67	12.8	1213	2 S16356	DNA-binding protei
29	66.5	12.7	146	2 H72167	ovo protein - frui
30					A35R protein - var

30 66.5 12.7 282 2 A71675  
31 66.5 12.7 539 2 B90595  
32 66.5 12.7 624 2 T28423  
33 66.5 12.7 678 2 S14173  
34 66.5 12.7 1238 2 A64596  
35 66 12.6 490 2 A46391  
36 66 12.6 2150 2 S71629  
37 65.5 12.5 306 2 T24402  
38 65.5 12.5 463 2 T31570  
39 65.5 12.5 2339 2 A45597  
40 65 12.5 398 1 R3BYM1  
41 65 12.5 618 2 S53384  
42 65 12.5 911 2 S28098  
43 64.5 12.4 200 2 G97048  
44 64.5 12.4 732 2 T23505  
45 64 12.3 273 2 A64238

#### ALIGNMENTS

##### RESULT 1

T09079

probable chloroquine resistance protein CG2 (strain 7G8) - malaria parasite (Plasm  
C:Species: Plasmodium falciparum  
C:Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 21-Jul-2000  
C:Accession: T09079  
R:Su, X.Z.; Kirkman, L.A.; Fujioaka, H.; Wellem, T.E.  
Cell 91, 593-603, 1997

A:Title: Complex polymorphisms in an 330 kDa protein are linked to Chloroquine-resi  
A:Reference number: Z16556; MUID:98054002; PMID:9393853  
A:Accession: T09079

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2708 <SUX>

A:Cross-references: EMBL:AF030692; NID:g2642513; PIDN:AAC47853.1; PID:g2642514

A:Experimental source: strain 7G8; from Brazil

C:Genetics:

A:Gene: cg2

C:Keywords: toxin resistance

Query Match 14.7%; Score 76.5; DB 2; Length 2708;  
Best Local Similarity 21.7%; Pred. No. 11;  
Matches 25; Conservative 24; Mismatches 37; Indels 29; Gaps 4;  
QY 2 WFKVGRKHQKFKVFWETD-----SNKTLVSLKKKPFHL-----YCVIYIP 45  
Db 631 WYKIG---YCVIFLFSILKIFCTDDTLLISRKHVYVLSDSLEKSEYSYKTVGIN 686  
QY 46 LVPKLIILFDIAFIPKSLISO-----FQNNHYTHNHTNHTNIRFNIIIS 91  
Db 687 IHDLLEDFDLHFIDKELKQDTNGNKYQDMNDHVKNDHTNYQNNHIFLLT 741

##### RESULT 2

T09080

probable chloroquine resistance protein CG2 (strain HB3) - malaria parasite (Plasm  
C:Species: Plasmodium falciparum  
C:Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 21-Jul-2000  
C:Accession: T09080

R:Su, X.Z.; Kirkman, L.A.; Fujioaka, H.; Wellem, T.E.  
Cell 91, 593-603, 1997

A:Title: Complex polymorphisms in an 330 kDa protein are linked to Chloroquine-resi

A:Reference number: Z16556; MUID:98054002; PMID:9393853

A:Accession: T09080

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2819 <SUX>

A:Cross-references: EMBL:AF030693; NID:g2642515; PIDN:AAC47854.1; PID:g2642516

A:Experimental source: strain HB3; from Honduras

C:Genetics:

A:Gene: cg2

C:Keywords: toxin resistance

100

---

## RESULT 7

## RESULT 7





Tue Apr 15 15:15:24 2003

us-09-830-244b-1.rpr

Page 5

OY 22 SNNKTLVSLKKKPFHLYCVIYIPLVPKLIILFDIAFIPKSLISQFQNNHYTHNH---T 78  
Db 1069 SNNYKFTIIRKKKKKKLCYIMKSFHPI----LDEFWNLSCQNEIKNIYYKNLHFVIS 1124  
OY 79 NHNTNNIRFNIIIS-----NCRT 95  
Db 1125 LHNSSIIDFKIINHFIILNKIFENISINCTT 1154

Search completed: April 14, 2003, 23:05:08  
Job time : 87 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2003, 22:34:36 ; Search time 84 Seconds  
(without alignments)  
46.908 Million cell updates/sec

Title: US-09-830-244B-1

Perfect score: 522

Sequence: 1 MFHFVKVRQHKFKVFWETD.....NHTNHTNNTNIRNFNIISNCR 95

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	68.5	13.1	499	1	UDPGL_YEAST	P32861	saccharomyc
2	68	13.0	346	1	DHAS_HELPY	O25801	rellicobacte
3	67	12.8	1028	1	OVO_DROME	P31521	droscophila
4	66.5	12.7	181	1	TBX3_MOUSE	P70324	mus musculus
5	66.5	12.7	678	1	SIR1_YEAST	P21691	saccharomyc
6	66	12.6	490	1	CAR2_DICDI	P02381	dictyosteli
7	65.5	12.5	396	1	RMAR_YEAST	P02381	saccharomyc
8	65.5	12.5	2339	1	RPC1_PLAFA	P27625	plasmodium
9	65	12.5	618	1	YJK3_YEAST	P42950	saccharomyc
10	65	12.5	757	1	ECR_LUCU	O18531	lucilia cup
11	65	12.5	826	1	CRAA_BACUH	Q9S597	bacillus th
12	65	12.5	911	1	STE6_SCHPO	P26674	schizosacch
13	64.5	12.4	544	1	PG10_DICDI	O06885	dictyosteli
14	64	12.3	273	1	ESL3_MYCCE	O49421	mycoplasma
15	64	12.3	375	1	CAR2_DICDI	P34907	dictyosteli
16	64	12.3	444	1	PR24_YEAST	P49960	saccharomyc
17	64	12.3	782	1	ZFY1_MOUSE	P10925	mus musculus
18	64	12.3	783	1	ZFY2_MOUSE	P20662	mus musculus
19	64	12.3	1177	1	SP97_DICDI	Q95293	dictyosteli
20	63.5	12.2	479	1	YK66_YEAST	O12194	saccharomyc
21	63	12.1	793	1	REGA_DICDI	Q23917	dictyosteli
22	63	12.1	1407	1	YCAA_DICDI	O03100	dictyosteli
23	62.5	12.0	490	1	PIT_BUCAI	P57847	buchnera ap
24	62.5	12.0	1570	1	P3K1_DICDI	P54673	dictyosteli
25	62	11.9	742	1	ZFA_MOUSE	P23607	mus musculus
26	62	11.9	969	1	DPMO_NEVIN	P33538	neurospora
27	61.5	11.8	516	1	YKE7_YEAST	P36090	saccharomyc
28	61.5	11.8	567	1	NU2M_HANWI	P48906	haemulana w
29	61.5	11.8	1901	1	YCF1_TOBAC	P12222	nicotiana t
30	61	11.7	237	1	YPV4_METTF	P29575	methanobact
31	61	11.7	317	1	AAC4_DICDI	P14198	dictyosteli
32	61	11.7	322	1	VF05_VARV	P33865	variola vir
33	61	11.7	448	1	AAC2_DICDI	P14196	dictyosteli

34 61 11.7 491 1 CAD3\_BOVIN  
35 61 11.7 658 1 REP\_BUCAP  
36 61 11.7 666 1 NX2B\_HUMAN  
37 61 11.7 1427 1 SRB8\_YEAST  
38 60.5 11.6 467 1 NORA\_CHLPN  
39 60.5 11.6 686 1 PTRB\_ECOLI  
40 60 11.5 148 1 Y722\_RICPR  
41 60 11.5 407 1 RP11\_YEAST  
42 60 11.5 490 1 CALR\_HUMAN  
43 60 11.5 554 1 YGS6\_YEAST  
44 60 11.5 603 1 HNIA\_XENLA  
45 60 11.5 666 1 YEAV\_YEAST

P19535 bos taurus  
O31889 buchnera ap  
P38401 homo sapien  
P25648 saccharomyc  
Q92792 chlamydia p  
P24555 escherichia  
Q92637 rickettsia  
P23250 saccharomyc  
P30908 homo sapien  
P46945 saccharomyc  
Q05041 xenopus lae  
P40002 saccharomyc

## ALIGNMENTS

RESULT 1  
UDPG\_YEAST  
ID UDPG\_YEAST STANDARD; PRT; 499 AA.  
AC P32861;  
DT 01-OCT-1993 (Rel. 27, Created)  
DT 01-OCT-1993 (Rel. 27, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE Probable UTP-glucose-1-phosphate uridylyltransferase (EC 2.7.7.9)  
DE (UDP-glucose pyrophosphorylase) (UDPGP) (UGPase).  
GN YKL035W OR YKL248.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288C;  
RA MEDLINE=93127731; PubMed=1481573;  
RA Purnelle B., Skala J., van Dyck L., Goffeau A.;  
RT "The sequence of a 12 kb fragment on the left arm of yeast chromosome  
RT XI reveals five new open reading frames, including a zinc finger  
RT protein and a homolog of the UDP-glucose pyrophosphorylase from  
RT potato.";  
RL Yeast 8:977-986(1992).  
CC -!- FUNCTION: PATHWAYS A CENTRAL ROLE AS A GLUCOSYL DONOR IN CELLULAR  
CC METABOLIC PATHWAYS (BY SIMILARITY).  
CC -!- CATALYTIC ACTIVITY: UTP + alpha-D-glucose 1-phosphate -  
CC diphosphate + UDP-glucose.  
CC -!- SIMILARITY: BELONGS TO THE EUKARYOTIC UDPGP FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X69584; CAA49303.1; -  
CC EMBL; 228035; CAA81872.1; -  
CC PIR; S30007; S30007.  
CC SGD; S0001518; YKL035W.  
CC InterPro; IPR002618; UDPGP.  
CC Pfam; PF01704; UDPGP; 1.  
KW Transferase; Kinase; Nucleotidyltransferase; Multigene family.  
SQ SEQUENCE 499 AA; 55988 MW; 9B9F4BED885E004D CRC64;  
-----

Query Match 13.1%; Score 68.5; DB 1; Length 499;  
Best Local Similarity 36.8%; Pred. No. 7.3;  
Matches 14; Conservative 8; Mismatches 13; Indels 3; Gaps 1;  
Oy 52 ILFLDIAFIPKSLISQFQNNHYTHNHTNNTNIRNFNI 89  
Db 291 VRLLEVAQVPRKEHIDFKN---IRKFTNFTNNTNINL 325













RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
RA Gellies S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
RA Mooney P., Moule S., Mungell K., Murphy L., Niblett D., Odell C.,  
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,  
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,  
RA Gabel C., Fuchs M., Fritze C., Holzer E., Moestl D., Hillbert H.,  
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Pohl T.M.,  
RA Borzym K., Langer I., Beck A., Lehrach H., Wambutt R., Pohl T.M.,  
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,  
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.,  
RT "The genome sequence of Schizosaccharomyces pombe.";  
RL Nature 415:871-880(2002).  
CC -!- FUNCTION: PROMOTES THE EXCHANGE OF RAS-BOUND GDP BY GTP. THIS  
CC PROTEIN IS ESSENTIAL FOR MATING.  
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 RAS-GEF DOMAIN.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X53254; CAA37345.1; -  
DR EMBL; AL049559; CAB40184.1; -  
DR EMBL; AL031966; CAA21435.1; -  
DR PIR; S28098; S28098.  
DR InterPro; IPR000851; RasGEFN.  
DR InterPro; IPR001895; RasGRF\_CDC25.  
DR InterPro; IPR001452; SH3.  
DR Pfam; PF00018; SH3; 1.  
DR Pfam; PF00617; RasGEF; 1.  
DR Pfam; PF00618; RasGEFN; 1.  
DR ProDom; PD000066; SH3; 1.  
DR SMART; SM00147; RasGEF; 1.  
DR SMART; SM00229; RasGEFN; 1.  
DR SMART; SM00326; SH3; 1.  
DR PROSITE; PS00720; GDS\_CDC25; 1.  
DR PROSITE; PS00002; SH3; 1.  
KW Guanine-nucleotide releasing factor; SH3 domain.  
FT DOMAIN 1 60  
SQ SEQUENCE 911 AA; 105185 MW; 5F8C12D20C4B753F CRC64;  
  
Query Match 12.5%; Score 65; DB 1; Length 911;  
Best Local Similarity 25.2%; Pred. No. 32;  
Matches 28; Conservative 10; Mismatches 23; Indels 50; Gaps 5;  
  
Qy 23 NKKTLVSL-----KKKKPHLY-----CVIYIPLVPLKLIILFLDIAFIPK----- 62  
Db 776 NSKTLASFELLNNLTARKNFNSYRDLNCV--LPCVPFLGVYFDLTFLTKGNKDNQ 833  
Qy 63 -----SLISQFNHHYTHNTN-----INTNNI 85  
Db 834 NAINFDKRTKVTIRLNIKKFQSVGYMFNPINEVQELLNEVISRERTNNI 884  
  
RESULT 13  
GP10\_DICDI  
ID GP10\_DICDI STANDARD; PRT; 544 AA.

AC Q05885;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE Glycoprotein GP100 precursor (P29F8).  
GN GPPA.  
OS Dictyostelium discoideum (Slime mold).  
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.  
OX NCBI\_TaxID=44689;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94124612; PubMed=8294503;  
RA Barth A., Mueller-Taubenberger A., Taranto P., Gerisch G.;  
RT "Replacement of the phospholipid-anchor in the contact site A  
RT glycoprotein of D. discoideum by a transmembrane region does not  
RT impede cell adhesion but reduces residence time on the cell  
RT surface.";  
RL J. Cell Biol. 124:205-215(1994).  
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -!- DEVELOPMENTAL STAGE: RESTRICTED TO THE AGGREGATION STAGE OF  
CC DEVELOPMENT IN D.DISCOIDEUM. NO DETECTABLE ACTIVITY IN CELL  
CC ADHESION.  
CC -!- FTM: N- AND O-GLYCOSYLATED.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; L04286; AAC37369.1; -  
DR DictyDb; D802035; gppa.  
KW Glycoprotein; Transmembrane; Signal.  
FT SIGNAL 1 19  
FT CHAIN 20 544  
FT DOMAIN 20 489  
FT TRANSMEM 490 510  
FT DOMAIN 511 544  
FT DOMAIN 117 208  
FT CARBOHYD 80 80  
FT CARBOHYD 224 224  
FT CARBOHYD 308 308  
FT CARBOHYD 332 332  
FT CARBOHYD 366 366  
FT CARBOHYD 380 380  
FT CARBOHYD 410 410  
FT CARBOHYD 422 422  
FT CARBOHYD 478 478  
SQ SEQUENCE 544 AA; 59156 MW; 9B2C2CDE7C7D0AA0 CRC64;  
  
Query Match 12.4%; Score 64.5; DB 1; Length 544;  
Best Local Similarity 26.9%; Pred. No. 21;  
Matches 21; Conservative 16; Mismatches 36; Indels 5; Gaps 2;  
  
Qy 18 ETDLSNNKTLVSLKKKPHLYCVIYIPLVPLKLIILFLDIAFIPKLSQFNHHYTHNH 77  
Db 472 DSSKSNNSGGGKNNK-----LYLLIILPTLVFLIIVAAVAIFI-KTRVSGNSGSKVKNKN 526  
Qy 78 TNHNTNNIRFNIIISNCR 95  
Db 527 NKKDSINYPFQMLDEITT 544  
  
RESULT 14  
ESL3\_MYCGE  
ID ESL3\_MYCGE STANDARD; PRT; 273 AA.  
AC Q49421; Q49335;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Putative esterase/lipase 3 (EC 3.1.-.-).





GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2003, 22:53:41 ; Search time 110 seconds  
(without alignments)  
177.950 Million cell updates/sec

Title: US-09-830-244B-1

Perfect score: 522

Sequence: 1 MFHFVKGRQHKFKVFWYD.....NFTNHTNNIRENIISNCKT 95

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: SPTREMBL\_21.\*
- 2: sp\_archaea.\*
- 3: sp\_bacteria.\*
- 4: sp\_fungi.\*
- 5: sp\_human.\*
- 6: sp\_invertebrate.\*
- 7: sp\_mammal.\*
- 8: sp\_mhc.\*
- 9: sp\_organelle.\*
- 10: sp\_phase.\*
- 11: sp\_plant.\*
- 12: sp\_rodent.\*
- 13: sp\_virus.\*
- 14: sp\_vertebrate.\*
- 15: sp\_unclassified.\*
- 16: sp\_virus.\*
- 17: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	94.5	18.1	512	8	Q9XMS7		Q9XMS7 tetrahymena
2	84	16.1	135	15	O90814		O90814 human immun
3	76.5	14.7	2708	5	O15791		O15791 plasmodium
4	76.5	14.7	2742	5	O15801		O15801 plasmodium
5	76.5	14.7	2819	5	O15792		O15792 plasmodium
6	76	14.6	215	10	Q9SIT9		Q9SIT9 arabidopsis
7	76	14.6	573	5	O77334		O77334 plasmodium
8	76	14.6	1672	5	Q8T2M6		Q8T2M6 dictyosteli
9	75	14.4	163	5	Q10054		Q10054 caenorhabdi
10	74.5	14.3	721	5	Q8T218		Q8T218 dictyosteli
11	74.5	14.3	2344	5	Q8SSV9		Q8SSV9 dictyosteli
12	74	14.2	141	12	Q90F67		Q90F67 vaccinia vi
13	73	14.0	426	16	O84141		O84141 chlamydia t
14	73	14.0	1317	5	Q8T4K2		Q8T4K2 dictyosteli
15	72.5	13.9	616	5	Q8TID3		Q8TID3 dictyosteli
16	72.5	13.9	2567	5	Q9U0J6		Q9U0J6 plasmodium

17	71.5	13.7	207	10	Q9AVX6		O9AVX6 guillardia
18	71.5	13.7	518	8	Q95029		Q95029 tetrahymena
19	71	13.6	215	2	Q93UA8		Q93UA8 carsonella
20	71	13.6	332	5	Q9XUN3		Q9XUN3 caenorhabdi
21	71	13.6	1146	3	O77362		O77362 plasmodium
22	71	13.6	1274	3	Q06673		Q06673 saccharomyc
23	70.5	13.5	1264	5	Q8WPV5		Q8WPV5 schistosoma
24	70.5	13.5	3432	5	Q95PH9		Q95PH9 dictyosteli
25	70	13.4	134	12	Q9EMW6		Q9EMW6 amsacta moo
26	69.5	13.3	154	15	Q9E8H5		Q9E8H5 human immun
27	69.5	13.3	343	12	Q91EW7		Q91EW7 cydia pomon
28	69.5	13.3	446	10	Q98RQ9		Q98RQ9 guillardia
29	69.5	13.3	548	5	O8TIS2		O8TIS2 dictyosteli
30	69.5	13.3	1017	5	O77371		O77371 plasmodium
31	69.5	13.3	1384	12	Q9EMR9		Q9EMR9 amsacta moo
32	69.5	13.3	2067	5	Q9U3U8		Q9U3U8 plasmodium
33	69.5	13.3	3417	5	O8T116		O8T116 dictyosteli
34	69	13.2	207	8	Q9G8W1		Q9G8W1 rhodomonas
35	69	13.2	454	5	O45096		O45096 caenorhabdi
36	69	13.2	605	5	O45170		O45170 caenorhabdi
37	69	13.2	1245	5	O96195		O96195 plasmodium
38	69	13.2	1281	5	O15802		O15802 plasmodium
39	68.5	13.1	215	2	Q93099		Q93099 carsonella
40	68.5	13.1	232	10	Q98S16		Q98S16 guillardia
41	68.5	13.1	502	3	Q9P966		Q9P966 candida gla
42	68.5	13.1	643	5	Q8ST22		Q8ST22 dictyosteli
43	68.5	13.1	743	5	Q9NGX0		Q9NGX0 dictyosteli
44	68.5	13.1	1560	5	Q9GRX5		Q9GRX5 dictyosteli
45	68.5	13.1	1561	5	Q9U987		Q9U987 dictyosteli

#### ALIGNMENTS

#### RESULT 1

Q9XMS7	Q9XMS7	PRELIMINARY;	PRT;	512 AA.
AC	Q9XMS7			
DT	01-NOV-1999	(TREMBLrel. 12, Created)		
DT	01-DEC-1999	(TREMBLrel. 12, Last sequence update)		
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)		
DE	Haem lyase.			
GN	YEJR.			
OS	Tetrahymena. pyriformis.			
OG	Mitochondrion.			
OC	Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;			
OC	Tetrahymena; Tetrahymena.			
OX	NCBI_TaxID=5908;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE-20181866; Pubmed-10715208;			
RA	Edqvist J., Burger G., Gray M.W.;			
RT	"Expression of mitochondrial protein-coding genes in Tetrahymena			
RT	pyriformis.";			
RL	J. Mol. Biol. 297:381-393(2000).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE-20181865; Pubmed-10715207;			
RA	Burger G., Zhu Y., Littlejohn T.G., Greenwood S.J., Schnare M.N.,			
RA	Lang B.F., Gray M.W.;			
RT	"Complete sequence of the mitochondrial genome of Tetrahymena			
RT	pyriformis and comparison with Paramacium aurelia mitochondrial DNA.";			
RL	J. Mol. Biol. 297:365-380(2000).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RA	Burger G.;			
RL	Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AF160864; AAD41936.1;			
DR	InterPro; IPR003567; Cyl_c_blog.			
DR	PRINTS; PR01410; CCBIOGENESIS.			
KW	Lyase; Mitochondrion.			
SK	SEQUENCE. 512 AA; 62577 MW; 6A4EF0808AC26366 CRC64;			





```

Db 643 WYKIG-----YCVFFLSILKIFCTCDDTTLLISKRHHVYLSDSLEKSEYSYKGTGVIN 698
Qy 46 LVPKLIILFDIAIPKSLISQ-----FQNNHYTHNTHNTNNTNIRFNIS 91
Db 699 IHDLDDYFFDLHFIDKELACQDGTNGIKFQDMNDHVKNDDHTNYQNNIFNLIT 753

RESULT 5
O15792 ID O15792 PRELIMINARY; PRT; 2819 AA.
AC O15792;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
GN CG2.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HB3;
RX MEDLINE=98054002; PubMed=99393853;
RA Su X., Kirkman L.A., Fujioka H., Wellem T.E.;
RT "Complex polymorphisms in an approximately kDa protein are linked to
RT chloroquine-resistant P. falciparum in Southeast Asia and Africa.";
RL Cell 91:593-603(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=HB3;
RX MEDLINE=20446201; PubMed=10989140;
RA Fidock D.A., Nomura T., Cooper R.A., Su X.-z., Talley A.K.,
RA Wellem T.E.;
RT "Allelic modifications of the cg2 and cg1 genes do not alter the
RT chloroquine response of drug-resistant Plasmodium falciparum.";
RL Mol. Biochem. Parasitol. 110:1-10(2000).
DR EMBL; AF030693; AAC47854.1;
SQ SEQUENCE 2819 AA; 335122 MW; FBF3CCCCOAC788D9 CRC64;

Query Match 14.78; Score 76.5; DB 5; Length 2819;
Best Local Similarity 21.78; Pred. No. 13;
Matches 25; Conservative 24; Mismatches 37; Indels 29; Gaps 4;

Qy 2 WFKVGRKHQKVFVETDL-----SNKTLVSLKPKPHL-----YCVIYIP 45
Db 637 WYKIG-----YCVFFLSILKIFCTCDDTTLLISKRHHVYLSDSLEKSEYSYKGTGVIN 692
Qy 46 LVPKLIILFDIAIPKSLISQ-----FQNNHYTHNTHNTNNTNIRFNIS 91
Db 699 IHDLDDYFFDLHFIDKELACQDGTNGIKFQDMNDHVKNDDHTNYQNNIFNLIT 747

RESULT 6
Q9SIT9 ID Q9SIT9 PRELIMINARY; PRT; 215 AA.
AC Q9SIT9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Putative CCAT-box-binding transcription factor.
GN AT2G13570.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae;
OC eusoids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.T., Feldblyum T.V.,

```

```

RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT thaliana.";
RL Nature 402:761-768(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Lin X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC007063; AAD22680.1;
DR HSSP; P48781; 1b67.
DR InterPro; IPR003956; CBFA_NFYB.
DR InterPro; IPR003958; CBFA_NFYB_domain.
DR InterPro; IPR003957; CBFA_NFYB_topis.
DR Pfam; PF00808; CBFD_NFYB_HMF.
DR PRINTS; PR00615; CCAATSUBUNTA.
DR PROSITE; PS00685; CBFA_NFYB.
SQ SEQUENCE 215 AA; 24619 MW; 0615B2B7A097DD25 CRC64;

Query Match 14.68; Score 76; DB 10; Length 215;
Best Local Similarity 27.08; Pred. No. 1.3;
Matches 24; Conservative 14; Mismatches 27; Indels 24; Gaps 3;

Qy 5 KVGKQHKVFVETDLNNKTLVSLKPKPHLYCVIYI-----PLVPKLIILFDIA 58
Db 145 QIQQNHNYQFOEQDQNNN-----NMCTSYISHHSPFLPVDHPQPNIA 192
Qy 59 FIPKSLISQFQNNHYTHNTHNTNNTNIRF 87
Db 193 FSPKSLQKQFPQOH-----DNNIDSIH 215

RESULT 7
O77334 ID O77334 PRELIMINARY; PRT; 573 AA.
AC O77334;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Putative dual-specificity protein phosphatase.
GN MAL3p3.11.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=99376085; PubMed=10448855;
RA Bowman S., Lawson D., Basham D., Brown D., Chillingworth T.,
RA Churcher C.M., Craig A., Davies R.M., Devlin K., Feltwell T.,
RA Gentles S., Gwilliam R., Hamlin N., Harris D., Holroyd S., Hornsby T.,
RA Horrocks P., Jagels K., Jassal B., Kyes S., McLean J., Moule S.,
RA Mungall K., Murphy L., Oliver K., Quail M.A., Rajadream M.-A.,
RA Rutter S., Skelton J., Squares R., Squares S., Sulston J.E.,
RA Whitehead S., Woodward J.R., Newbold C., Barrell B.G.;
RT "The complete nucleotide sequence of chromosome 3 of Plasmodium
RT falciparum.";
RL Nature 400:532-538(1999).
DR EMBL; Z98547; CAB11119.2;
DR InterPro; IPR000340; DS_phosphatase.
DR InterPro; IPR000387; TYR_phosphatase.
DR Pfam; PF00782; DSPC; 1.
DR SMART; SM00012; PTPC_DSPC; 1.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; UNKNOWN_1.
DR PROSITE; PS00056; TYR_PHOSPHATASE_2; 2.
DR PROSITE; PS00054; TYR_PHOSPHATASE_DUAL; 2.
SQ SEQUENCE 573 AA; 67573 MW; 97E2AUAFA37D3334 CRC64;

```

Query Match 14.6%; Score 76; DB 5; Length 573;  
Best Local Similarity 26.4%; Pred. No. 3.4;  
Matches 23; Conservative 17; Mismatches 33; Indels 14; Gaps 3;  
QY 8 RKOHEKVTWEVDLSNNKTLVSLKKKPFHLYCVIYIPVLPKLIILFDIAFIPKSLI-- 65  
Db 132 RKEHYDIIFPSDIINNTCCNNKLD-----YI-----KSMILKEDAYIDFDVIHM 179  
QY 66 SQFQNNHHTHNTNHNIRNFIISN 92  
Db 180 DQKKNHNNNNNNNNNNNNNNNNNNNN 206  
RESULT 8  
Q8T2M6 PRELIMINARY; PRT; 1672 AA.  
AC Q8T2M6;  
DT 01-JUN-2002 (TReMBLrel. 21, Created)  
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)  
DE Hypothetical 185.4 kDa protein.  
OS Dictyostelium discoideum (Slime mold).  
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.  
OX NCBI\_TaxID=44689;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AX4;  
RA Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,  
RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,  
RA Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;  
RT "Sequence and Analysis of Chromosome 2 of Dictyostelium";  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AAL15593; AAL92294.1; -;  
KW Hypothetical protein.  
SQ SEQUENCE 1672 AA; 185413 MW; BEBE0A8644989875 CRC64;

Query Match 14.6%; Score 76; DB 5; Length 1672;  
Best Local Similarity 27.7%; Pred. No. 9.1;  
Matches 23; Conservative 12; Mismatches 24; Indels 24; Gaps 3;  
QY 19 TDLNNKTLVSLK-----KKKPFHLYCVIYIPVLPKLIILFDIAFIPKSLISQFQ 69  
Db 225 TEIPNLSTLNLTKIESNFSSEAKIYCYLYCOL-----PFDL-SNNN 269  
QY 70 NNHYTHNHTNHNIRNFIISN 92  
Db 270 NNNNNNNNNNNNNNNNNNNNNNN 292

RESULT 9  
Q10054 PRELIMINARY; PRT; 163 AA.  
ID Q10054  
AC Q10054;  
DT 01-NOV-1996 (TReMBLrel. 01, Created)  
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
DE F09C12.6 protein.  
GN F09C12.6.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Pelodierinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX MEDLINE=94150718; PubMed=7906398;  
RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M.,  
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkneen R.,

RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,  
RA Thierly-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
elegans.";  
RL Nature 368:32-38(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Anderson K.;  
RL "The sequence of C. elegans cosmid F09C12.";  
RN Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Waterston R.;  
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U28929; AAA68347.1; -;  
SQ SEQUENCE 163 AA; 18602 MW; 2305A1F3E7290CB5 CRC64;  
Query Match 14.4%; Score 75; DB 5; Length 163;  
Best Local Similarity 25.2%; Pred. No. 1.3;  
Matches 26; Conservative 20; Mismatches 37; Indels 20; Gaps 4;  
QY 1 MFHFKVGRKHQHFVTFWETDLSNNKTLVSLKKKPFHLYCVIYIPVLPKLIILFDIAFI 60  
Db 63 IWFHKNGRNR-----FENNDAIVAGIKFLRSSSEFASTLVLLTLLIGPV-IAYR 111  
QY 61 PKSL-----ISQFQNNHHTHNTNHNIRNFIISNCR 95  
Db 112 KRNKVLVYCGKKAAGNDYATFIYI-CNISGYLFSFLNCIT 153  
RESULT 10  
Q8T218 PRELIMINARY; PRT; 721 AA.  
ID Q8T218  
AC Q8T218;  
DT 01-JUN-2002 (TReMBLrel. 21, Created)  
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)  
DE Hypothetical 83.4 kDa protein.  
OS Dictyostelium discoideum (Slime mold).  
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.  
OX NCBI\_TaxID=44689;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AX4;  
RA Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,  
RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,  
RA Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;  
RT "Sequence and Analysis of Chromosome 2 of Dictyostelium";  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC116030; AAL92987.1; -;  
KW Hypothetical protein.  
SQ SEQUENCE 721 AA; 83384 MW; A7A8C83BE1FE5BC1 CRC64;  
Query Match 14.3%; Score 74.5; DB 5; Length 721;  
Best Local Similarity 29.1%; Pred. No. 6.1;  
Matches 30; Conservative 18; Mismatches 30; Indels 25; Gaps 7;  
QY 11 HFK---VTWE---TDL---SNNKTLVSLKKKPFHLYC-----VIYIPVLPKLIILFD 56  
Db 149 HFKKGVMLSEKIFLDLFYNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 201  
QY 57 IAFIPKSLISQF--QNNHY--THNHTNHNHTNHNIRNFIISNCR 95  
Db 202 ISFIHSHLFDYLNNN 244  
RESULT 11  
Q8SSV9 PRELIMINARY; PRT; 2344 AA.  
ID Q8SSV9  
AC Q8SSV9;

```
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Putative transferase.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA Tunngal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RA "Sequence and Analysis of Chromosome 2 of Dictyostellium.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC115598; AAL93230.1; -
SQ SEQUENCE 2344 AA; 266642 MW; A897406B0A4634 CRC64;

Query Match
Best Local Similarity 14.3%; Score 74.5; DB 5; Length 2344;
Matches 28; Conservative 15; Mismatches 28; Indels 31; Gaps 6;

QY 2 WFHKVGRKHQKVFETDLSNNKTLVSLKK-KP-----FHLVCYIYIPLV 47
Db 1697 WFTKC-----HFETI-----NKTIRAFKQNKPTSPDVLHDLIAEFQYRLISLS 1744
QY 48 ---PKLIILFLDIAFIPKSLISQFQNN--HYTHNHTNNTNN 84
Db 1745 SVNPSALFRPTITQPGFDLNYFNTHHHHHHHHHHGN 1786

RESULT 12
Q9JF67 PRELIMINARY; PRT; 141 AA.
AC Q9JF67;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DE T4A0R.
OS Vaccinia virus (strain Tian Tan).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
OX NCBI_TaxID=10253;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TIAN TAN;
RA Jin Q., Hou Y.D., Cheng N.H., Yao E.M., Cheng S.X., Yang X.K.,
RA Jing D.Y., Yu W.H., Yuan J.S., Ma X.J.;
RA "Complete genomic sequence of vaccinia virus (Tian Tan strain).";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF095689; AAF34036.1; -
DR InterPro: IPR000504; RNA_rec_mot.
DR PROSITE: PS00030; RRM_RNP.1; UNKNOWN.1.
SQ SEQUENCE 141 AA; 16504 MW; 3B0DF51BDF9A0231 CRC64;

Query Match
Best Local Similarity 14.2%; Score 74; DB 12; Length 141;
Matches 23; Conservative 27; Mismatches 30; Indels 44; Gaps 4;

QY 13 KVTFWETDLSNNKTLVSLKKK-----KPFHLCYIYI-----PLVP----- 48
Db 13 KTSFVNCNDSTREKIKIKHKGMSFVYKPKHSTVYKLGSGGIYHDDLVLGKVTINDL 72
QY 49 KLIILFLDIAF-----IPKSLISQFQNNHYTHNHTNNTNIREN 88
Db 73 KMLFYMDSLHGVTSSGVYIKLGSSTDRLSLNRITVTKVNNYNNYNNYNNYNN 132
QY 89 IISN 92
Db 133 NYNN 136

RESULT 13
```

```
084141
ID 084141 PRELIMINARY; PRT; 426 AA.
AC 084141;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE Oligopeptide binding protein.
GN OPPA_1 OR Ctl39.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UM-3/CX;
RX MEDLINE=99000809; PubMed=9784136;
RA Stephens R.S., Kaiman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.;
RA Davis R.W.;
RA "Genome sequence of an obligate intracellular pathogen of humans:
RA Chlamydia trachomatis.";
RL Science 282:754-759(1998).
DR EMBL: AE001287; AAC67730.1; -
DR InterPro: IPR000914; SBP_bac_5.
DR Pfam: PF00496; SBP_bac_5; 2.
KW Complete proteome.
SQ SEQUENCE 426 AA; 48271 MW; 8C1FF71717980235 CRC64;

Query Match
Best Local Similarity 14.0%; Score 73; DB 16; Length 426;
Matches 30; Conservative 19; Mismatches 34; Indels 30; Gaps 7;

QY 2 WFH--KVGR-----KQHEKVFETDLSNNKTLVSLKKKPFHL-----YCVIYIPLV 47
Db 109 WEHTKAGRYSLLEKLSFRAS-----SSSELLHELKEPEQLLAILASPFAVYRPN 162
QY 48 PKLIILFLDIA-FIPKSLIS----QFQNNHYTHNHTNNTNIRENITSNCR 95
Db 163 P-----FLSSGPFMPKTYVQGTVLQKNPYVDHAEHLHSDIRIIPNIT 210

RESULT 14
Q8T4K2 PRELIMINARY; PRT; 1317 AA.
ID Q8T4K2;
AC Q8T4K2;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE ABC transporter ABCC14.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Anjard C., Loomis W.F.;
RA "Evolution of the ABC transporter of Dictyostellium.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY082376; AAL89812.1; -
SQ SEQUENCE 1317 AA; 150238 MW; 1768BBFD23B24D1A CRC64;

Query Match
Best Local Similarity 14.0%; Score 73; DB 5; Length 1317;
Matches 31; Conservative 16; Mismatches 39; Indels 44; Gaps 5;

QY 8 RKQHFK-----VTFWETD-----LSNNKTLVSLKKKPFHL-----YCVIY 43
Db 330 RKQELKFLFLVFLWFIWDHMIETNATLVVSTFATVLTGNTMSLDVTFAMTIFANLK 389
QY 44 IPLVPKLIILFLDIAFIP-----KSLISQFQNNHYTHNHTNNTNIREN--- 88
Db 390 LPLIYLPEDIYKAIGLMPVKRIQNFLLKSSLESKRENNQINFNNNNNNNNNNN 449
QY 89 -----IISNC 93
```

Search completed: April 14, 2003, 23:03:32  
Job time : 114 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: April 14, 2003, 23:06:51 ; Search time 147.7 Seconds  
(without alignments)  
1043.807 Million cell updates/sec

Title: US-09-830-244B-1  
Perfect score: 522  
Sequence: 1 MWFKVGRKQHKVTFWETD.....NHTNHTNIRNFNIISNCR 95

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlp  
-Q/Cgn2\_1/USPFC.spool/US09830244/runat\_10042003\_090016\_27874/app.query.fasta\_1.263  
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40 cdi-LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09830244.CGN\_1.1.2874 @runat\_10042003\_090016\_27874 -NCPU=6 -ICPU=3  
-NO\_XLPY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPOP=0.5 -FGAPOP=6 -FGAPOP=7  
-YGAPOP=10 -YGAPOP=0.5 -DELOP=6 -DELEXT=7

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_esthu:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_fod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
c 1	349	66.9	306	10	BE163886	BE163886 QV3-HT046
c 2	321	61.5	951	12	BE168510	BE168510 602339663
c 3	294	56.3	329	14	BQ308679	BQ308679 RC6-BT070
c 4	201	38.5	615	12	BF678862	BF678862 602153173
c 5	180.5	34.6	286	9	AA376683	AA376683 EST89136
c 6	178.5	34.2	304	9	AA376720	AA376720 EST89178
c 7	178.5	34.2	341	9	AA376695	AA376695 EST89151
c 8	178.5	34.2	426	9	AA376680	AA376680 EST89133
c 9	177	33.9	321	12	BG205925	BG205925 RST25360
c 10	176	33.7	335	9	AA376714	AA376714 EST89171
c 11	176	33.7	470	12	BG200817	BG200817 RST20025
c 12	174.5	33.4	517	12	BE188342	BE188342 RST7360 A
c 13	173.5	33.2	519	10	AW951480	AW951480 EST363550
c 14	171.5	32.9	457	9	AA376682	AA376682 EST89135
c 15	164	31.4	404	9	AA376713	AA376713 EST89170
c 16	156	29.9	405	9	AA376719	AA376719 EST89177
c 17	150	28.7	375	9	AA376681	AA376681 EST89134
c 18	148	28.4	714	10	BE253236	BE253236 601116852
c 19	145	27.8	410	14	T29650	T29650 EST89132 Hu
c 20	144.5	27.7	514	9	AA195115	AA195115 ZR35B03.r
c 21	144	27.6	334	9	AA744925	AA744925 np72e22.s
c 22	144	27.6	373	10	BE464523	BE464523 hv84h04.x
c 23	144	27.6	440	9	AI421621	AI421621 tf38g10.x
c 24	144	27.6	449	9	AA747799	AA747799 nx86a06.s
c 25	144	27.6	458	9	AI150564	AI150564 qf36e06.x
c 26	144	27.6	461	9	AI377872	AI377872 te63h11.x
c 27	144	27.6	477	13	BM663533	BM663533 UT-E-CO-
c 28	144	27.6	495	14	BQ427792	BQ427792 AGENCOURT
c 29	144	27.6	496	10	AW444891	AW444891 UI-H-B13-
c 30	144	27.6	519	14	BM956861	BM956861 1j27d11.y
c 31	144	27.5	559	14	BM936624	BM936624 1j27d11.y
c 32	143.5	27.5	373	12	BF477953	BF477953 7f08a06.x
c 33	143.5	27.5	394	12	BF115130	BF115130 h74q05.x
c 34	143.5	27.5	427	12	BF477850	BF477850 7f06a06.x
c 35	143.5	27.5	428	10	AW613151	AW613151 hh42a06.x
c 36	143.5	27.5	451	10	AW189807	AW189807 x109c07.x
c 37	143.5	27.5	469	10	AW293206	AW293206 UI-H-B12-
c 38	143.5	27.5	513	9	AA195114	AA195114 z35a04.f
c 39	143.5	27.5	613	10	AV727718	AV727718 AV727718
c 40	143	27.4	321	10	AW001596	AW001596 wr99g09.x
c 41	143	27.4	440	10	AW468605	AW468605 he42b05.x
c 42	143	27.4	619	9	AA961420	AA961420 OF53a12.s
c 43	142.5	27.3	521	10	BE465101	BE465101 hv75b07.x
c 44	142	27.2	374	10	BE247773	BE247773 TCBA2P553
c 45	141	27.0	373	9	AI692383	AI692383 wd63g06.x

## ALIGNMENTS

RESULT 1  
LOCUS BE163886/c 306 bp mRNA linear EST 21-JUN-2000  
DEFINITION QV3-HT0461-230200-103-h02 HT0461 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BE163886  
VERSION BE163886.1 GI:8626607  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 306)  
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,



Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J., and Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

Proc Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

## TITLE

## JOURNAL

## MEDLINE

## COMMENT

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL

(<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC6&t2=RC6-BT0709-310300-021-F02&t3=2000-03-31&t4=1>)

Seq primer: puc 18 forward

High quality sequence start: 22

High quality sequence stop: 57

Location/Qualifiers

1. 329

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="BT0709"

/dev\_stage="Adult"

/note="Organ: breast; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 106 a 53 c 60 g 110 t

ORIGIN

Alignment Scores:

Pred. No.: 7 93e-26 Length: 329

Score: 294.00 Matches: 60

Percent Similarity: 96.83% Conservative: 1

Best Local Similarity: 95.24% Mismatches: 0

Query Match: 56.32% Indels: 2

DB: 14 Gaps: 0

US-09-830-244B-1 (1-95) x BQ309679 (1-329)

Qy 6 ValGlyArgLysGlnHisPheLysValThrPheTrpGluThrAspLeuSerAsnLys 25

Db 324 GTGGGAAGAAACAGCATTTTAAAGTAACCTTTTGGGAGACTGATTGAGTAATAATAA 265

Qy 26 ThrLeuValSerLeuLys--LysLysLysProPheHisLeuTyrCysValIleTyrIleP 45

Db 264 ACTCTGCTCCCTTAGAAGAAAAAACCCTTCCACCTTACTGTGCAITATATACC 205

Qy 45 roLeuValProLysLeuIlelePheLeuAspIleAlaPheIleProLysSerLeuI 65

Db 204 CCTAGTCCCAAGTAATATCTTTATTTCTGGATATTGCTTTATACCAAGACCTTAA 145

Qy 65 leSer 66

Db 144 TCAGC 140

RESULT 4

BF678862

LOCUS

DEFINITION

602153173Fl NIH\_MGC\_83 Homo sapiens cDNA clone IMAGE:4294170 5', mRNA sequence.

ACCESSION

BF678862

VERSION

BF678862.1 GI:11952757

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

EST.

human.

Homo sapiens

1 (Bases 1 to 615)

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

NIH-MGC <http://mgc.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: [cgabs-r@mail.nih.gov](mailto:cgabs-r@mail.nih.gov)

Tissue Procurement: CLONETECH Laboratories, Inc.

cDNA Library Preparation: CLONETECH Laboratories, Inc.

DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)

Clone distribution: Incyte Genomics, Inc.

Found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLCML143 row: a column: 19

High quality sequence stop: 569.

Location/Qualifiers

1. 615

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="NIH\_MGC\_83"

/lab\_host="DH10B (T1 phage-resistant)"

/note="Organ: prostate; Vector: pDNR-LIB (Clontech); Site\_1: SfiI (ggcgctcgcc); Site\_2: SfiI (ggccattatggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CAGGCCATTAGGCG-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGGCGGCGGACATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.4 kb (range 0.5-4.0 kb). 14/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

BASE COUNT 150 a 128 c 173 g 164 t

ORIGIN

Alignment Scores:

Pred. No.: 2 87e-14 Length: 615

Score: 201.00 Matches: 47

Percent Similarity: 96.00% Conservative: 1

Best Local Similarity: 94.00% Mismatches: 2

Query Match: 38.51% Indels: 2

DB: 12 Gaps: 0

US-09-830-244B-1 (1-95) x BF678862 (1-615)

Qy 17 TrpGluThrAspLeuSerAsnLysThrLeuValSerLeuLysLysLysProphe 36

Db 1 TGGGAGACTGATTG-AGTAATAATAAACTCTGCTCCCTTAAGAAAAAACCCCTTC 59

Qy 37 HisLeuTyrCysValIleTyrIleProLeuValProLysLeuIleLeuPheLeuAsp 56

Db 60 CACCTTTACTGTGCTTATATATCCCTTAGTCCAAAGTAAATATCTATTCTCGAT 119

Qy 57 IleAlaPheIleProLysSerLeuIleSer 66

Db 120 AT-CCTTTTATACCAAGACCCCTTATCAGC 148

RESULT 5

AA376683

LOCUS

DEFINITION

EST89136 Salivary gland Homo sapiens cDNA 5' end similar to statherin, mRNA sequence.

ACCESSION

AA376683

VERSION

AA376683.1 GI:2029001

KEYWORDS

EST.

human.

Homo sapiens

286 bp mRNA linear EST 21-APR-1997

AA376683

LOCUS

DEFINITION

EST89136 Salivary gland Homo sapiens cDNA 5' end similar to statherin, mRNA sequence.

ACCESSION

AA376683

VERSION

AA376683.1 GI:2029001

KEYWORDS

EST.

human.

Homo sapiens

286 bp mRNA linear EST 21-APR-1997

AA376683

LOCUS

DEFINITION

EST89136 Salivary gland Homo sapiens cDNA 5' end similar to statherin, mRNA sequence.

ACCESSION

AA376683

VERSION

AA376683.1 GI:2029001

KEYWORDS

EST.

human.

Homo sapiens

286 bp mRNA linear EST 21-APR-1997

AA376683

LOCUS

DEFINITION

EST89136 Salivary gland Homo sapiens cDNA 5' end similar to statherin, mRNA sequence.

ACCESSION

AA376683

VERSION

AA376683.1 GI:2029001

KEYWORDS

EST.

human.

Homo sapiens

286 bp mRNA linear EST 21-APR-1997

AA376683

LOCUS

DEFINITION

EST89136 Salivary gland Homo sapiens cDNA 5' end similar to statherin, mRNA sequence.

ACCESSION

AA376683

VERSION

AA376683.1 GI:2029001

KEYWORDS

EST.

human.

Homo sapiens

286 bp mRNA linear EST 21-APR-1997

AA376683

LOCUS

DEFINITION

EST89136 Salivary gland Homo sapiens cDNA 5' end similar to statherin, mRNA sequence.

ACCESSION

AA376683

VERSION

AA376683.1 GI:2029001

KEYWORDS

EST.

human.

Homo sapiens

286 bp mRNA linear EST 21-APR-1997

AA376683

LOCUS

DEFINITION

EST89136 Salivary gland Homo sapiens cDNA 5' end similar to statherin, mRNA sequence.

ACCESSION

AA376683

VERSION

AA376683.1 GI:2029001

KEYWORDS

EST.

human.

Homo sapiens

286 bp mRNA linear EST 21-APR-1997

AA376683

LOCUS

DEFINITION

EST89136 Salivary gland Homo sapiens cDNA 5' end similar to statherin, mRNA sequence.

ACCESSION

AA376683

VERSION

AA376683.1 GI:2029001

KEYWORDS

EST.

human.

Homo sapiens

286 bp mRNA linear EST 21-APR-1997

AA376683

LOCUS

DEFINITION

EST89136 Salivary gland Homo sapiens cDNA 5' end similar to statherin, mRNA sequence.

ACCESSION

AA376683

VERSION

AA376683.1 GI:2029001

KEYWORDS

EST.

human.

Homo sapiens

286 bp mRNA linear EST 21-APR-1997

AA376683

LOCUS

DEFINITION

EST89136 Salivary gland Homo sapiens cDNA 5' end similar to statherin, mRNA sequence.

ACCESSION

AA376683

VERSION

AA376683.1 GI:2029001

KEYWORDS

EST.

human.

Homo sapiens

286 bp mRNA linear EST 21-APR-1997

AA376683

LOCUS

DEFINITION

EST89136 Salivary gland Homo sapiens cDNA 5' end similar to statherin, mRNA sequence.

ACCESSION

AA376683

VERSION

AA376683.1 GI:2029001

KEYWORDS

EST.

human.

Homo sapiens

286 bp mRNA linear EST 21-APR-1997

AA376683

LOCUS

DEFINITION

EST89136 Salivary gland Homo sapiens cDNA 5' end similar to statherin, mRNA sequence.

ACCESSION



REFERENCE 1 (bases 1 to 286)  
 AUTHORS Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A., Bult, C.J., Lee, N.H., Kirkness, E.F., Weinstein, K.G., Gocayne, J.D., White, O., Sutton, G., Blake, J.A., Brandon, R.C., Man-Wai, C., Clayton, R.A., Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D., Fitzgerald, L.M., Fitzhugh, W.M., Fritchman, J.L., Geoghagen, N.S., Glodek, A., Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr., Kelley, J.M., Kelley, J.C., Liu, L.-I., Marmaros, S.M., Merrick, J.M., Moreno-Palauques, R.F., McDonald, L.A., Nguyen, D.T., Pelligrino, S.M., Phillips, C.A., Ryder, S.E., Scott, J.L., Saudak, D.M., Shirley, R., Small, K.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y., Bednarek, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J., Dimke, D., Feng, D.-F., Ferrie, A., Fischer, C., Hastings, G.A., He, W.W., Kunsch, C., Hungjun, J., Li, H., Weissner, P.S., Olsen, H., Raymond, L., Wei, Y.F., Wing, J., Xu, C., Yu, G.L., Ruben, S.M., Dillion, P.J., Fannon, M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M. and Venter, J.C.

TITLE Initial assessment of human gene diversity and expression patterns  
 JOURNAL based upon 83 million nucleotides of cDNA sequence  
 MEDLINE Nature 377 (6547 Suppl), 3-174 (1995)  
 COMMENT Other ESTs: EST89134 EST89135 THC77891  
 Contact: Kerlavage, AR  
 Bioinformatics  
 The Institute for Genomic Research  
 9712 Medical Center Drive, Rockville, MD 20850 USA  
 Tel: 3018699056  
 Fax: 3018699423  
 Email: arkerlav@tigr.org  
 For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (<http://www.tigr.org/tdb/hgi/hgi.html>)  
 Seq primer: M13 Reverse.

FEATURES  
 source  
 1. 286  
 /organism="Homo sapiens"  
 /db\_xref="ATCC (inhost):18111"  
 /db\_xref="taxon:9606"  
 /clone\_lib="Salivary gland"  
 /dev\_stage="adult"  
 /note="Organ: salivary gland; Vector: pBluescript SK-";  
 Site\_1: ECORI; Site\_2: XhoI  
 BASE COUNT 82 a 67 c 48 g 88 t 1 others  
 ORIGIN

Alignment Scores:  
 Pred. No.: 3.49e-12 Length: 286  
 Score: 180.50 Matches: 40  
 Percent Similarity: 60.76% Conservative: 8  
 Best Local Similarity: 50.63% Mismatches: 18  
 Query Match: 34.58% Indels: 13  
 DB: 9 Gaps: 2  
 US-09-830-244B-1 (1-95) x AA376683 (1-286)  
 Qy 30 LeuLysLysLysLysProPheHisLeuTyrcysValleTyrllePro----- 45  
 Db 31 TTGAGTAAAGAACACCCACCACTATGAGTTCCTTCCTTCCTTCATCTGGGTC 90  
 Qy 46 ---LeuValProLysLysLysLysLeuLeuLeuPheLeuAspIleAlaPhe----- 59  
 Db 91 TCATGGTTTCATGATGGGCTGATTCATCTGAAGAGAAATTTTCGCTAGAAATGGA 150  
 Qy 60 -----TleProLysSerLeuLeuSerGlnPheGlnAsnHisTyrrThrHisAsn 76  
 Db 151 GATTCGGTTATGGGTATGGCTTTATCAGCCAGTTCCAGAACCAACCACTATACCAAC 210  
 Qy 77 HisThrAsnHisAsnThrAsnAsnIleArgPheAsnIleSerAsnCysArgThr 95  
 Db 211 CATACCAACCAACATACCAACATATACCTTTTATATCATCATGTAACCTGCAGGACA 267

RESULT 6  
 AA376720  
 LOCUS  
 DEFINITION AA376720 Salivary gland Homo sapiens cDNA 5' end similar to statherin, mRNA sequence.  
 ACCESSION AA376720  
 VERSION AA376720.1 GI:2029038  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 304)  
 AUTHORS Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A., Bult, C.J., Lee, N.H., Kirkness, E.F., Weinstein, K.G., Gocayne, J.D., White, O., Sutton, G., Blake, J.A., Brandon, R.C., Man-Wai, C., Clayton, R.A., Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D., Fitzgerald, L.M., Fitzhugh, W.M., Fritchman, J.L., Geoghagen, N.S., Glodek, A., Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr., Kelley, J.M., Kelley, J.C., Liu, L.-I., Marmaros, S.M., Merrick, J.M., Moreno-Palauques, R.F., McDonald, L.A., Nguyen, D.T., Pelligrino, S.M., Phillips, C.A., Ryder, S.E., Scott, J.L., Saudak, D.M., Shirley, R., Small, K.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y., Bednarek, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J., Dimke, D., Feng, D.-F., Ferrie, A., Fischer, C., Hastings, G.A., He, W.W., Kunsch, C., Hungjun, J., Li, H., Weissner, P.S., Olsen, H., Raymond, L., Wei, Y.F., Wing, J., Xu, C., Yu, G.L., Ruben, S.M., Dillion, P.J., Fannon, M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M. and Venter, J.C.

TITLE Initial assessment of human gene diversity and expression patterns  
 JOURNAL based upon 83 million nucleotides of cDNA sequence  
 MEDLINE Nature 377 (6547 Suppl), 3-174 (1995)  
 COMMENT Other ESTs: EST89177 THC77891  
 Contact: Kerlavage, AR  
 Bioinformatics  
 The Institute for Genomic Research  
 9712 Medical Center Drive, Rockville, MD 20850 USA  
 Tel: 3018699056  
 Fax: 3018699423  
 Email: arkerlav@tigr.org  
 For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (<http://www.tigr.org/tdb/hgi/hgi.html>)  
 Seq primer: M13 Reverse.

FEATURES  
 source  
 1. 304  
 /organism="Homo sapiens"  
 /db\_xref="ATCC (inhost):181127"  
 /db\_xref="taxon:9606"  
 /clone\_lib="Salivary gland"  
 /dev\_stage="adult"  
 /note="Organ: salivary gland; Vector: pBluescript SK-";  
 Site\_1: ECORI; Site\_2: XhoI  
 BASE COUNT 89 a 72 c 50 g 93 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 6.54e-12 Length: 304  
 Score: 178.50 Matches: 37  
 Percent Similarity: 63.77% Conservative: 7  
 Best Local Similarity: 53.62% Mismatches: 8  
 Query Match: 34.20% Indels: 17  
 DB: 9 Gaps: 2  
 US-09-830-244B-1 (1-95) x AA376720 (1-304)  
 Qy 27 LeuValSerLeuLysLysLysLysProPheHisLeuTyrcysValleTyrlleProLeu 46  
 Db 88 CTGATCATCTGAAGAGAAAT-----TTTTCGCTA----- 117  
 Qy 47 ValProLysLeuIleLeuPheLeuAspIleAlaPheIleProLysSerLeuIleSer 66





```

/clone_lib="Salivary gland"
/dev_stage="adult"
/notes="Organ: salivary gland; Vector: pBluescript SK-;
Site_1: EcoRI; Site_2: XhoI"
BASE COUNT      128 a   97 c   64 g   135 t   2 others
ORIGIN

Alignment Scores:
Pred. No.:      9.74e-12      Length:      426
Score:          178.50      Matches:      37
Percent Similarity: 63.77%      Conservatives: 7
Best Local Similarity: 53.62%      Mismatches: 8
Query Match:      34.20%      Indels:      17
DB:              9          Gaps:      2

US-09-830-244B-1 (1-95) x AA376680 (1-426)

Qy 27 LeuValSerLeuLysLysLysProPheHisLeuTyrCysValIleTyrIleProLeu 46
Db 116 CTGATTCATCTGAAGAGAAAT-----TTTTCGGTA----- 145
Qy 47 ValProLysLeuIleLeuPheLeuAspIleAlaPheIleProLysSerLeuIleSer 66
Db 146 -----GAATTCGAAGATTCGGTTATGGGTATGGCCCTTATCAGC 184
Qy 67 GlnPheGlnAsnHisTyrThrHisAsnHisThrAsnHisAsnThrAsnAsnIleArg 86
Db 185 CAGTTCAGACACCACTATACCCACCAACCAATACCAACCAATACCAATATACCT 244
Qy 87 PheAsnIleLeuSerAsnCysArgThr 95
Db 245 TTTAATATCATCATGTAAGTACTGAGGACA 271

RESULT 9
Bg205925
LOCUS      Bg205925      321 bp      mRNA      linear      EST 21-APR-2001
DEFINITION R5125360 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION  Bg205925
VERSION    Bg205925.1 GI:13727612
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE  1 (bases 1 to 321)
AUTHORS   Cain, S., Leventhal, C., Thornton, M., Ramchandran, R., Whittington, J.,
Lerner, L., Costanzo, D., McElligott, K., Boozer, S., Mays, R., Smith
E., Veloso, N., Klika, A., Hess, J., Cothren, K., Lo, K., Offenbacher
J., Danzig, J. and Ducar, M.
Creation of genome-wide protein expression libraries using random
activation of gene expression
Nat. Biotechnol. 19 (5), 440-445 (2001)
21227151
Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scain@atersys.com
High quality sequence stop: 271.
Location/Qualifiers
1. 321
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/cell_line="HT1080"
/notes="See 'Creation of Genome-wide Protein Expression
Libraries using Random Activation of Gene Expression',
Nature Biotechnology, in press. Note that even though the
cell type indicated is HT1080, since a random activation
method was used, these sequence tags are not necessarily
expressed in HT1080 under normal circumstances."

FEATURES
source
1. 321

```

```

BASE COUNT      97 a   78 c   48 g   95 t   3 others
ORIGIN

Alignment Scores:
Pred. No.:      1.06e-11      Length:      321
Score:          177.00      Matches:      32
Percent Similarity: 87.80%      Conservatives: 4
Best Local Similarity: 78.05%      Mismatches: 5
Query Match:      33.91%      Indels:      0
DB:              12          Gaps:      0

US-09-830-244B-1 (1-95) x BG205925 (1-321)

Qy 55 LeuAspIleAlaPheIleProLysSerLeuIleSerGlnPheGlnAsnAsnHisTyrThr 74
Db 125 TTGGAGATTCGGTTATGGGTATGGCCCTTATCAGCAGTTCACCAACCACTATACC 184
Qy 75 HisAsnHisThrAsnHisAsnThrAsnAsnIleArgPheAsnIleIleSerAsnCysArg 94
Db 185 CACAACCATACCAACCAATACCAATATACCTTTTAAATATCATCATGTAAGTGCAGG 244
Qy 95 Thr 95
Db 245 ACA 247

RESULT 10
AA376714
LOCUS      AA376714      335 bp      mRNA      linear      EST 21-APR-1997
DEFINITION EST89171 Salivary gland Homo sapiens cDNA 5' end similar to
statherin, mRNA sequence.
ACCESSION  AA376714
VERSION    AA376714.1 GI:2029032
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE  1 (bases 1 to 335)
AUTHORS   Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A., Bult
C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D., White
O., Sutton, G., Blake, J.A., Brandon, R.C., Man-Wei, C., Clayton, R.A.,
Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D., Fitzgerald
L.M., Fitchugh, W.M., Fritchman, J.B., Geoghagen, N.S., Glodek, A.,
Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr., Kelley, J.M.,
Kelley, J.C., Liu, L.-I., Martiros, S.M., Merrick, J.M.,
Moreno-Palauques, R.F., McDonald, L.A., Nguyen, D.T., Peilgrino, S.M.,
Phillips, C.A., Ryder, S.E., Scott, J.L., Saudek, D.M., Shirley, R.,
Small, K.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y.,
Bednarik, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J.,
Dimke, D., Feng, D.-F., Ferrie, A., Fischer, C., Hastings, G.A., He, W.W.,
Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kozak, D.L.,
Kunsch, C., Hungjun, J., Li, H., Meissner, P.S., Olsen, H., Raymond, L.,
Wei, Y.F., Wing, J., Xu, C., Yu, G.L., Ruben, S.M., Dillon, P.J., Fannon
M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M. and
Venter, J.C.
Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
Nature 377 (6547 Suppl), 3-174 (1995)
96026280
Other ESTs: EST89170 THC77891
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlavet@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (http://www.tigr.org/tdb/hgi/hgi.html)
Seq primer: M13 Reverse.
Location/Qualifiers
1. 335

```

```

/organism="Homo sapiens"
/db_xref="ATCC (inhost):181124"
/db_xref="taxon:9606"
/clone_lib="Salivary gland"
/dev_stage="adult"
/note="Organ: salivary gland; Vector: pBluescript SK-;
Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 109 a 75 c 43 g 107 t 1 others
ORIGIN

Alignment Scores:
Pred. No.: 1.47e-11 Length: 335
Score: 176.00 Matches: 31
Percent Similarity: 96.97% Conservative: 1
Best Local Similarity: 93.94% Mismatches: 1
Query Match: 33.72% Indels: 0
DB: 9 Gaps: 0

US-09-830-244B-1 (1-95) x AA376714 (1-335)

QY 63 SerLeuileSerGlnPheGlnAsnHsIstYrThrHisAsnHsThrAsnHsAsnThr 82
Db 21 GCCCTTATCAGCCAGTTCACAGAACCACTATACCCAGAACCAATACCAATACC 80
QY 83 AsnAsnIleArgPheAsnIleIleSerAsnCysArgThr 95
Db 81 AACAATATACCTTTTATATATCATCATGTAAGTGCAGGACA 119

RESULT 11
LOCUS BG200817 470 bp mRNA linear EST 21-APR-2001
DEFINITION RST20025 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION BG200817
VERSION BG200817.1 GI:13722392
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 470)
AUTHORS Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
Cain,S., Leventhal,C., Thornton,M., Ramachandran,R., Whittington,J.,
Lerner,L., Costanzo,D., McElligott,K., Booser,S., Mays,R., Smith
,J., Veloso,N., Klika,A., Hess,J., Cothren,K., Lo,K., Offenbacher
,J., Danzig,J. and Ducar,M.
Creation of genome-wide protein expression libraries using random
activation of gene expression
Nat. Biotechnol. 19 (5), 440-445 (2001)
21227151
COMMENT Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scaine@atersys.com
High quality sequence stop: 470.
Location/Qualifiers
1..470
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/cell_line="HT1080"
/note="See 'Creation of Genome-wide Protein Expression
Libraries using Random Activation of Gene Expression',
Nature Biotechnology, in press. Note that even though the
cell type indicated is HT1080, since a random activation
method was used, these sequence tags are not necessarily
expressed in HT1080 under normal circumstances."
BASE COUNT 156 a 99 c 61 g 154 t
ORIGIN

Alignment Scores:
Pred. No.: 2.19e-11 Length: 470
Score: 174.50 Matches: 36
Percent Similarity: 62.32% Conservative: 7
Best Local Similarity: 52.11% Mismatches: 9
Query Match: 33.43% Indels: 17
DB: 12 Gaps: 2

US-09-830-244B-1 (1-95) x BG188342 (1-517)

QY 27 LeuValSerLeuLysLysLysLysPropheHisLeuTyrcysValIleTyrlleProLeu 46
Db 74 CTGATTTCATCTGAAGAGAAAT-----TTTTCGCTA----- 103

```

```

Score: 176.00 Matches: 31
Percent Similarity: 96.97% Conservative: 1
Best Local Similarity: 93.94% Mismatches: 1
Query Match: 33.72% Indels: 0
DB: 12 Gaps: 0

US-09-830-244B-1 (1-95) x BG200817 (1-470)

QY 63 SerLeuileSerGlnPheGlnAsnHsIstYrThrHisAsnHsThrAsnHsAsnThr 82
Db 83 GCCCTTATCAGCCAGTTCACAGAACCACTATACCCAGAACCAATACCAATACC 142
QY 83 AsnAsnIleArgPheAsnIleIleSerAsnCysArgThr 95
Db 143 AACATATACCTTTTATATATCATCATGTAAGTGCAGGACA 181

RESULT 12
LOCUS BG188342 517 bp mRNA linear EST 21-APR-2001
DEFINITION RST7360 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION BG188342
VERSION BG188342.1 GI:13710029
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 517)
AUTHORS Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
Cain,S., Leventhal,C., Thornton,M., Ramachandran,R., Whittington,J.,
Lerner,L., Costanzo,D., McElligott,K., Booser,S., Mays,R., Smith
,J., Veloso,N., Klika,A., Hess,J., Cothren,K., Lo,K., Offenbacher
,J., Danzig,J. and Ducar,M.
Creation of genome-wide protein expression libraries using random
activation of gene expression
Nat. Biotechnol. 19 (5), 440-445 (2001)
21227151
COMMENT Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scaine@atersys.com
High quality sequence stop: 517.
Location/Qualifiers
1..517
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/cell_line="HT1080"
/note="See 'Creation of Genome-wide Protein Expression
Libraries using Random Activation of Gene Expression',
Nature Biotechnology, in press. Note that even though the
cell type indicated is HT1080, since a random activation
method was used, these sequence tags are not necessarily
expressed in HT1080 under normal circumstances."
BASE COUNT 171 a 110 c 73 g 163 t
ORIGIN

Alignment Scores:
Pred. No.: 3.72e-11 Length: 517
Score: 174.50 Matches: 36
Percent Similarity: 62.32% Conservative: 7
Best Local Similarity: 52.11% Mismatches: 9
Query Match: 33.43% Indels: 17
DB: 12 Gaps: 2

US-09-830-244B-1 (1-95) x BG188342 (1-517)

QY 27 LeuValSerLeuLysLysLysLysPropheHisLeuTyrcysValIleTyrlleProLeu 46
Db 74 CTGATTTCATCTGAAGAGAAAT-----TTTTCGCTA----- 103

```

```

QY 47 ValProLysLeuLeileLeuPheLeuAspIleAlaPheLeuProLysSerLeuIleSer 66
Db 104 -----GAATTGGAAGATTTCGGTTATGGGTATGGCCCTTATCAGC 142
QY 67 GlnPheGlnAsnHisThrHisAsnHisThrAsnHisAsnThrAsnAsnIleArg 86
Db 143 CAGTTCAGAACCAACCATATACCAACCAACCAACCAACCAACCAACCAATATACCT 202
QY 87 PheAsnIleIleSerAsnCysArgThr 95
Db 203 TTTAATATCTTCAGTAAGTTCAGGACA 229

RESULT 13
LOCUS AW951480 519 bp mRNA linear EST 01-JUN-2000
DEFINITION EST263550 MAGE resequences, MAGB Homo sapiens cDNA, mRNA sequence.
ACCESSION AW951480
VERSION AW951480.1 GI:8141155
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 519)
AUTHORS Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt
,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
Quackenbush,J.
TITLE Assessment of gene expression patterns in a model of colon tumor
metastasis using a 19,200 element cDNA microarray
JOURNAL Unpublished (2000)
COMMENT Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johnq@tigr.org
Plate: 27
Seq primer: Reverse.
FEATURES
source Location/Qualifiers
1..519
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MAGE resequences, MAGB"
/note="Vector: pBluescriptSKm"
BASE COUNT 168 a 113 c 79 g 159 t
ORIGIN
Alignment Scores:
Pred. No.: 4.94e-11 Length: 519
Score: 173.50 Matches: 36
Percent Similarity: 63.77% Conservative: 8
Best Local Similarity: 52.17% Mismatches: 8
Query Match: 33.24% Indels: 17
DB: 10 Gaps: 2
US-09-830-244B-1 (1-95) x AW951480 (1-519)
QY 27 LeuValSerLeuLysLysLysLysProPheHisLeuTyrCysValIleTyrIleProLeu 46
Db 111 CTGATTCATCTGAGAGAAAT-----TTTTCGCTA----- 140
QY 47 ValProLysLeuLeileLeuPheLeuAspIleAlaPheIleProLysSerLeuIleSer 66
Db 141 -----GAATTGGAAGATTTCGGTTATGGGTATGGCCCTTATCAGC 179
QY 67 GlnPheGlnAsnHisThrHisAsnHisThrAsnHisAsnThrAsnAsnIleArg 86
Db 180 CAGTTCAGAACCAACCATATACCAACCAACCAACCAACCAACCAACCAATATACCT 239
QY 87 PheAsnIleIleSerAsnCysArgThr 95
Db 240 TTTAATATCTTCAGTAAGTTCAGGACA 266

```

## RESULT 14

```

AA376682
LOCUS AA376682 457 bp mRNA linear EST 21-APR-1997
DEFINITION EST9135 Salivary gland Homo sapiens cDNA 5' end similar to
statherin, mRNA sequence.
ACCESSION AA376682
VERSION AA376682.1 GI:2029000
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 457)
AUTHORS Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White
,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,
Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodzik,A.,
Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr., Kelley,J.M.,
Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W.,
Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,
Kunsch,C., HungJun,J., Li,H., Melissner,P.S., Olsen,H., Raymond,L.,
Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon
,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
Venter,J.C.
TITLE Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
JOURNAL Nature 377 (6547 Suppl), 3-174 (1995)
MEDLINE 96026280
COMMENT Other ESTs: EST89134 EST89136 THC77891
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (http://www.tigr.org/tdb/hgi/hgi.html)
Seq primer: M13 Reverse.
FEATURES
source Location/Qualifiers
1..457
/organism="Homo sapiens"
/db_xref="ATCC (inhost):181111"
/db_xref="taxon:9606"
/clone_lib="Salivary gland"
/dev_stage="adult"
/note="Organ: salivary gland; Vector: pBluescript SK-;
Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 133 a 100 c 70 g 150 t
ORIGIN
Alignment Scores:
Pred. No.: 7.42e-11 Length: 457
Score: 171.50 Matches: 36
Percent Similarity: 62.32% Conservative: 7
Best Local Similarity: 52.17% Mismatches: 9
Query Match: 32.85% Indels: 17
DB: 9 Gaps: 2
US-09-830-244B-1 (1-95) x AA376682 (1-457)
QY 27 LeuValSerLeuLysLysLysLysProPheHisLeuTyrCysValIleTyrIleProLeu 46
Db 102 CTGATTCATCTGAGAGAAAT-----TTTTCGCTA----- 131
QY 47 ValProLysLeuLeileLeuPheLeuAspIleAlaPheIleProLysSerLeuIleSer 66

```

```

Db 132 -----GAATTGGAAGATTCGGTTATGGTATGGCCCTTATCAGC 170
QY 67 GlnPheGlnAsnAsnHisTyrThrHisAsnHisThrAsnHisAsnHisAsnThr 82
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 171 CAGTCCAGACACACCATATACCCACACCAATACCAACAATATACCT 230
QY 87 PheAsnIleIleSerAsnCysArgThr 95
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 231 TTTTANTATCATCAGTAACAGCAGACA 257

RESULT 15
AA376713/c
LOCUS AA376713
DEFINITION EST89170 Salivary gland Homo sapiens cDNA 3' end similar to
statherin, mRNA sequence.
ACCESSION AA376713
VERSION AA376713.1 GI:2029031
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 404)
AUTHORS Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White
,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,
Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A.,
Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr., Kelley,J.M.,
Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
Moreno-Palantes,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Sauder,D.M., Shirley,R.,
Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
Dimke,D., Feng,D.-F., Ferlie,A., Fischer,C., Hastings,G.A., He,W.W.,
Kunsch,C., Hunglun,J., Li,H., Meissner,P.S., Olsen,H., Kozak,D.L.,
Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon
,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
Venter,J.C.
Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
Nature 377 (6547 Suppl), 3-174 (1995)
96026280
Other_ESTs: EST89171 THC77891
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (http://www.tigr.org/tdb/hgi/hgi.html)
Seq primer: M13-21.
Location/Qualifiers
1. .404
/organism="Homo sapiens"
/db_xref="ATCC (inhost):181124"
/db_xref="taxon:9606"
/clone_lib="Salivary gland"
/dev_stage="adult"
/note="Organ: salivary gland; Vector: pBluescript SK-;
Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 123 a 54 c 88 g 133 t 6 others
ORIGIN

```

```

Best Local Similarity: 87.88% Mismatches: 2
Query Match: 31.42% Indels: 0
DB: 9 Gaps: 0
US-09-830-244B-1 (1-95) x AA376713 (1-404)
QY 63 SerLeuIleSerGlnPheGlnAsnAsnHisTyrThrHisAsnHisThrAsnHisAsnThr 82
: : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 376 GCCCTTATCAGCCAGGTCAGGACACCACTATACCCACCAATACCAACAATACCT 278
QY 83 AsnAsnIleArgPheAsnIleIleSerAsnCysArgThr 95
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 316 AACAAATATACCCNTTTTAAATATATCATCAGTAACAGCAGACA 278

Search completed: April 15, 2003, 00:02:02
Job time : 1479 secs

```

Alignment Scores:  
Pred. No.: 5.17e-10 Length: 404  
Score: 164.00 Matches: 29  
Percent Similarity: 93.94% Conservative: 2

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 14, 2003, 20:01:06 ; Search time 3567 Seconds  
(without alignments)  
10859.499 Million cell updates/sec

Title: US-09-830-244B-2  
Perfect score: 1331  
Sequence: 1 ctatgttttagaatcaag.....gcctttccaatgtcacttg 1331

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues  
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_ov.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sy.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vl.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_fun.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_mu.\*
- 21: em\_or.\*
- 22: em\_or.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sy.\*
- 28: em\_sy.\*
- 29: em\_vl.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_inv.\*
- 33: em\_htg\_inv.\*
- 34: em\_htg\_inv.\*
- 35: em\_htg\_inv.\*
- 36: em\_htg\_inv.\*
- 37: em\_htg\_inv.\*
- 38: em\_htg\_inv.\*
- 39: em\_htg\_inv.\*
- 40: em\_htg\_inv.\*
- 41: em\_htg\_inv.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	704	52.9	133984	9	AL512306	AL512306 Human DNA
C 2	396	29.8	153023	2	AC021462	AC021462 Homo sapi
C 3	361.2	27.1	153023	2	AC021462	AC021462 Homo sapi
C 4	349.2	26.2	146805	2	AL450424	AL450424 Homo sapi
C 5	253	19.0	542	9	HUMSTRNA	HUMSTRNA
C 6	253	19.0	552	9	HUMSTRNA	HUMSTRNA
C 7	253	19.0	1584	9	AK092678	AK092678 Homo sapi
C 8	176.2	13.2	297	11	G05711	G05711 human SRS S
C 9	145.4	10.9	1498	11	G26643	G26643 human SRS S
C 10	145.4	10.9	4723	9	HUMSTRNA2	HUMSTRNA2
C 11	145.4	10.9	141568	9	AC063956	AC063956 Homo sapi
C 12	145.4	10.9	161549	2	AC024676	AC024676 Homo sapi
C 13	86.2	6.5	524	9	HUMSTRNA	HUMSTRNA
C 14	86.2	6.5	558	9	HUMSTRNA	HUMSTRNA
C 15	81	6.1	480	9	HUMSTRNA	HUMSTRNA
C 16	81	6.1	566	9	HUMSTRNA	HUMSTRNA
C 17	78.8	5.9	491	9	HUMSTRNA	HUMSTRNA
C 18	77.2	5.8	209410	2	AL805933	AL805933 Mus muscu
C 19	74.6	5.6	9881	6	AX281311	AX281311 Sequence
C 20	72	5.4	193445	2	AL845502	AL845502 Mus muscu
C 21	64.8	4.9	81962	2	AC128396	AC128396 Rattus no
C 22	64.8	4.9	246173	2	AC094026	AC094026 Rattus no
C 23	62.2	4.7	185969	2	AC069037	AC069037 Homo sapi
C 24	59	4.4	1764	9	HUMSTRNA	HUMSTRNA
C 25	55.8	4.2	9881	6	AX281312	AX281312 Sequence
C 26	52.4	3.9	164399	3	PFMAL3P6	PFMAL3P6
C 27	52.2	3.9	128635	9	AC108074	AC108074 Homo sapi
C 28	52.2	3.9	154472	2	AC109457	AC109457 Homo sapi
C 29	52.2	3.9	207558	2	AC114778	AC114778 Homo sapi
C 30	51.6	3.9	170141	9	AC092022	AC092022 Homo sapi
C 31	51.6	3.9	171044	2	AC092787	AC092787 Homo sapi
C 32	51.6	3.9	194874	2	AC080090	AC080090 Homo sapi
C 33	51.6	3.9	274349	2	AC093623	AC093623 Homo sapi
C 34	51.2	3.8	12029	3	AE001373	AE001373 Plasmodiu
C 35	51	3.8	182362	9	AL160033	AL160033 Human DNA
C 36	48.6	3.7	1141	6	AX083744	AX083744 Sequence
C 37	48.6	3.7	157051	9	AC011095	AC011095 Homo sapi
C 38	48.6	3.7	167830	2	AC027768	AC027768 Homo sapi
C 39	48.6	3.7	173765	9	AC090457	AC090457 Homo sapi
C 40	48.4	3.6	6056	6	AX345928	AX345928 Sequence
C 41	48.4	3.6	121709	9	AC093783	AC093783 Homo sapi
C 42	48.4	3.6	129989	9	AC005681	AC005681 Homo sapi
C 43	48.4	3.6	174277	9	AC093826	AC093826 Homo sapi
C 44	48	3.6	61215	2	AC090357	AC090357 Homo sapi
C 45	48	3.6	179934	9	AP000810	AP000810 Homo sapi

ALIGNMENTS

RESULT 1  
AL512306/c  
LOCUS  
DEFINITION Human DNA sequence from clone RP11-430C7 on chromosome 1, complete  
ACCESSION AL512306  
VERSION AL512306.16 GI:18491332  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 133984)  
AUTHORS Bray-Allen,S.  
TITLE Direct Submission

## JOURNAL

Submitted (01-FEB-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

On Feb 4, 2002 this sequence version requested: clonerequest@sanger.ac.uk  
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at

[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/chr1>  
Rp11-430C7 is from the library RPCR-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>  
VECTOR: pBACE3.6

IMPORTANT: This sequence is not the entire insert of clone Rp11-430C7. It may be shorter because we sequence overlapping sections only once, except for a short overlap. The true right end of clone Rp11-430C7 is at 133984 in this sequence. The true left end of clone Rp11-739N20 is at 97715 in this sequence. The true right end of clone Rp11-2317 is at 2000 in this sequence.

## FEATURES

source Location/Qualifiers  
1..133984  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/clone="Rp11-430C7"  
/clone\_lib="RPCR-11.2"  
123904..123933  
misc\_feature  
/note="single clone region. Reads generated from a transposon library derived from a single pUC clone. Restriction digest data confirm the assembly."  
BASE COUNT 37368 a 32097 c 30202 g 34317 t  
ORIGIN  
Query Match 52.9%; Score 704; DB 9; Length 133984;  
Best Local Similarity 99.2%; Pred. No. 1.6e-140;  
Matches 718; Conservative 0; Mismatches 5; Indels 1; Gaps 1;  
QY 1 CTATGTTTTAGAAATCAAGATGACCGGTAAAGCTGTCTCATGTACCAACGCTGAATTT 60  
DB 92638 CTATGTTTTAGAAATCAAGATGACCGGTAAAGCTGTCTCATGTACCAACGCTGAATTT 92579  
QY 61 ACAGTGTTCACAAATGCTGGAAATTTGTCATGCGCCATAGGGAATCTTAAGTTTACTTGGC 120  
DB 92578 ACAGTGTTCACAAATGCTGGAAATTTGTCATGCGCCATAGGGAATCTTAAGTTTACTTGGC 92519  
QY 121 TGAATTTATCAGCTTGTGACTAACAAGTTGAGTTTACAGATGAGGGGGAATATTG 180  
DB 92518 TGAATTTATCAGCTTGTGACTAACAAGTTGAGTTTACAGATGAGGGGGAATATTG 92459  
QY 181 AGGCCCTTAAGCTAAACAAATAATCAGTATCTCAGATAGTGGCTAATGTCCTCCCA 240  
DB 92458 AGGCCCTTAAGCTAAACAAATAATCAGTATCTCAGATAGTGGCTAATGTCCTCCCA 92399  
QY 241 GGCCTAATTTGGGAACAGTTTTCTGATGCTTTTGAAGTACTTTCTTTTGACAGAAA 300  
|||||

DB 92398 GGCCTAATTTGGGAACAGTTTTTCTGATTCCTTGTGAGAACTACTTCTTTTGACAGAAA 92339  
QY 301 TTTTCATTTCTGCTTCCCATTTGCTATATTTCTCCCTTTATAGAGCAATGGATTTCTTTCC 360  
DB 92338 TTTTCATTTCTGCTTCCCATTTGCTATATTTCTCCCTTTATAGAGCAATGGATTTCTTTCC 92279  
QY 361 TTTTCGCGAAATGTCCTCATTTAGCATTTTTCAGATCTTTTGTGATGTCGCACTAATGCAATAT 420  
DB 92278 TTTTCGCGAAATGTCCTCATTTAGCATTTTTCAGATCTTTTGTGATGTCGCACTAATGCAATAT 92219  
QY 421 TGTGTAATCCCTGTTATTTGGTGAATACAGCATAGTTTAAATAAATCTTTACAGTAAATCTTACA 480  
DB 92218 TGTGTAATCCCTGTTATTTGGTGAATACAGCATAGTTTAAATAAATCTTTACAGTAAATCTTACA 92159  
QY 481 CTGGAATTTGCTGCACCTCTACCAATAGCCCTTTTGAATGACTGAAGTGTGTAACAGAGAA 540  
DB 92158 CTGGAATTTGCTGCACCTCTACCAATAGCCCTTTTGAATGACTGAAGTGTGTAACAGAGAA 92099  
QY 541 AGAGCATGCTCTGCAGAAAGAGATAGCTAATATTTTGGTACTTTATCTGAAATCCCAAG 600  
DB 92098 AGAGCATGCTCTGCAGAAAGAGATAGCTAATATTTTGGTACTTTATCTGAAATCCCAAG 92039  
QY 601 ATGCTGCTTCCCTGCGAGGTTGTTTCTCTTACGATCCCTCAATGTAATCCCTCTCGGA 660  
DB 92038 ATGCTGCTTCCCTGCGAGGTTGTTTCTCTTACGATCCCTCAATGTAATCCCTCTCGGA 91979  
QY 661 GCACGAGCAGTATAGTAGAATCTCCATTTCTTTGTTTCTTTTAAAGACAGAGACTCT 720  
DB 91978 GCACGAGCAGTATAGTAGAATCTCCATTTCTTTGTTTCTTTTAAAGACAGAGACTCT 720  
QY 721 GTCT 724  
|||  
DB 91919 CTCT 91916

## RESULT 2

AC021462/c

LOCUS

DEFINITION

Homo sapiens clone RP11-276C1, WORKING DRAFT SEQUENCE, 21 unordered

pieces.

AC021462

VERSION

AC021462.3 GI:7387343

KEYWORDS

HTG; HTGS\_PHASE1; HTGS\_DRAFT.

SOURCE

Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 153023)

Birten,B., Linton,L., Nusbaum,C. and Lander,E.

Autors

TITLE

Unpublished

REFERENCE

2 (bases 1 to 153023)

Birten,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,

Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F., Allen,N.,

Boguslavsky,L., Boukhgaltier,B., Brown,A., Burkett,G., Castle,A.,

Chapel,I., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,

DeArrellano,K., Dewar,K., Domino,M., Doyle,M., Fensholt,J.,

Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,

Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,

Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,

Landers,T., Lehoczy,J., Levine,R., Lied,C., Liu,G., Locke,K.,

Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,

McPheeters,R., Meldrim,J., Menes,L., Morrow,J., Naylor,J.,

Norman,C.H., O'Connor,T., O'Donnell,P., Olivari,T.M., Peterson,K.,

Pierre,N., Pisanl,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,

Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,

Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,

Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,

Zimmer,A. and Zody,M.

Direct Submission

JOURNAL

Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT

On Apr 3, 2000 this sequence version replaced gi:7230200.



All repeats were identified using RepeatMasker:  
Smt, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WtBR  
Web site: <http://www.seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: L5201  
Center clone name: 276\_C\_1

```

----- Summary Statistics -----
Sequencing vector: M43; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 141815 bases at least Q40
Consensus quality: 147368 bases at least Q30
Consensus quality: 149481 bases at least Q20
Insert size: 157000; agarose-fp
Insert size: 151023; sum-of-contigs
Quality coverage: 4.3 in Q20 bases; agarose-fp
Quality coverage: 4.5 in Q20 bases; sum-of-contigs
-----

* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

```

```

source
1. .153023
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-276C1"
/clone_lib="RPC1-11 Human Male BAC"
1. .1015
/note="assembly_fragment"
1116. .2276
/note="assembly_fragment"
2377. .4009
/note="assembly_fragment"
4110. .5737
/note="assembly_fragment"
5838. .6939
/note="assembly_fragment"
7040. .9504
/note="assembly_fragment"
9605. .11734
/note="assembly_fragment"
11835. .14111
/note="assembly_fragment"
14212. .16973
/note="assembly_fragment"
17074. .19588
/note="assembly_fragment"
19689. .22275
/note="assembly_fragment"
22376. .25823
/note="assembly_fragment"
25924. .31307
/note="assembly_fragment"
31408. .34863
/note="assembly_fragment"
34964. .41382
/note="assembly_fragment"
41483. .49024
/note="assembly_fragment"
49125. .58672
/note="assembly_fragment"
58773. .69622
/note="assembly_fragment"
69723. .88191
/note="assembly_fragment"
88292. .107084
/note="assembly_fragment"
107185. .153023
/note="assembly_fragment
clone_end:T7
vector_side:left"
BASE COUNT 41494 a 35192 c 34573 g 39756 t 2008 others
ORIGIN
Query Match 29.8%; Score 396; DB 2; Length 153023;
Best Local Similarity 100.0%; Pred. No. 8.1e-75;
Matches 396; Conservative 0; Mismatches 0; Indels 0; Gaps
Qy 1 CTATGTTTTAGAAATCAAGATGAACGGTAGAGTGTCTCATGTACCAAACTGAAATTT 50
Db 4505 CTATGTTTTAGAAATCAAGATGAACGGTAGAGTGTCTCATGTACCAAACTGAAATTT 4446
Qy 61 ACAGTGTTTACAATGTCTGGAATTTTGCACATGCCATAGGGAATGTTAAGGTTACTTGGC 120
Db 4445 ACAGTGTTTACAATGTCTGGAATTTTGCACATGCCATAGGGAATGTTAAGGTTACTTGGC 4386
Qy 121 TGGAAATTATCAGACTTGTGAGTAACAAGTTGAAGTTTAGCAGATGAGGGGAATATTTG 180
Db 4385 TGGAAATTATCAGACTTGTGAGTAACAAGTTGAAGTTTAGCAGATGAGGGGAATATTTG 4326
Qy 181 AGGCCCTTAAGGCTAAACAAATATCATGATCTGAGATAGTGGCTATGTGGCTCCCCA 240
Db 4325 AGGCCCTTAAGGCTAAACAAATATCATGATCTGAGATAGTGGCTATGTGGCTCCCCA 4266
Qy 241 GGCCTAATTTGGGAACAGTTTTTTCCTGATTGCTTTTGAGAAGTACPTTCTTTTGACAGAA 300

```



```

Db 4265 GGCCTAAATTTGGGAACAGATTTTCTGATGCTTTGAGAAGTACATTTCTTTTGACAGAAA 4206
Qy 301 TTTTCATTTCTGTTCCCATTTGCTATATCTCCCTTTATAGAGCCATTTGGATTTCTTTCC 360
Db 4205 TTTTCATTTCTGTTCCCATTTGCTATATCTCCCTTTATAGAGCCATTTGGATTTCTTTCC 4146
Qy 361 TTTTGTGGGAATGTTCCCATTTAGCATTTTTCAGATCT 396
Db 4145 TTTTGTGGGAATGTTCCCATTTAGCATTTTTCAGATCT 4110

AC021462 153023 bp DNA linear HTG 03-APR-2000
LOCUS Homo sapiens clone RP11-276C1, WORKING DRAFT SEQUENCE, 21 unordered
pieces.
AC021462
AC021462.3 GI:7387343
HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 153023)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens, clone RP11-276C1
Unpublished
2 (bases 1 to 153023)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
Choepli,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
Dearellano,K., Dewar,K., Domino,M., Doyle,M., Fenesstor,J.,
Ferrell,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
McDonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheeters,R., Melidrim,J., Meneus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Olivari,T.M., Peterson,K.,
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A., and Zody,M.
Direct Submission
Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 3, 2000 this sequence version replaced gi:7230200.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html
Smit, A.F.A. & Green, P. (1996-1997)
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L5201
Center clone name: 276_C_1
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 141815 bases at least Q40
Consensus quality: 147368 bases at least Q30
Consensus quality: 149481 bases at least Q20
Insert size: 157000; agarose-fp
Insert size: 151023; sum-of-contigs
Quality coverage: 4.3 in Q20 bases; agarose-fp
Quality coverage: 4.5 in Q20 bases; sum-of-contigs
-----

```

\* NOTE: This is a 'working draft' sequence. It currently consists of 21 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

* 1 1015: contig of 1015 bp in length
* 1016 1115: gap of 100 bp
* 1116 2276: contig of 1161 bp in length
* 2277 2376: gap of 100 bp
* 2377 4009: contig of 1633 bp in length
* 4010 4109: gap of 100 bp
* 4110 5737: contig of 1628 bp in length
* 5738 5837: gap of 100 bp
* 5838 6939: contig of 1102 bp in length
* 6940 7039: gap of 100 bp
* 7040 9504: contig of 2465 bp in length
* 9505 9604: gap of 100 bp
* 9605 11734: contig of 2130 bp in length
* 11735 11834: gap of 100 bp
* 11835 14111: contig of 2277 bp in length
* 14112 14211: gap of 100 bp
* 14212 16973: contig of 2762 bp in length
* 16974 17073: gap of 100 bp
* 17074 19588: contig of 2515 bp in length
* 19589 19688: gap of 100 bp
* 19689 22275: contig of 2587 bp in length
* 22276 22375: gap of 100 bp
* 22376 25823: contig of 3448 bp in length
* 25824 25923: gap of 100 bp
* 25924 31307: contig of 5384 bp in length
* 31308 31407: gap of 100 bp
* 31408 34863: contig of 3456 bp in length
* 34864 34963: gap of 100 bp
* 34964 41382: contig of 6419 bp in length
* 41383 41482: gap of 100 bp
* 41483 49024: contig of 7542 bp in length
* 49025 49124: gap of 100 bp
* 49125 58672: contig of 9548 bp in length
* 58673 58772: gap of 100 bp
* 58773 69622: contig of 10850 bp in length
* 69623 69722: gap of 100 bp
* 69723 88191: contig of 18469 bp in length
* 88192 88291: gap of 100 bp
* 88292 107084: contig of 18793 bp in length
* 107085 107184: gap of 100 bp
* 107185 153023: contig of 45839 bp in length.

```

## FEATURES

## source

```

1..153023
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-276C1"
/clone_lib="RPC1-11 Human Male BAC"

misc_feature
1..1015
/note="assembly_fragment"

misc_feature
1116..2276
/note="assembly_fragment"

misc_feature
2377..4009
/note="assembly_fragment"

misc_feature
4110..5737
/note="assembly_fragment"

misc_feature
5838..6939
/note="assembly_fragment"

misc_feature
7040..9504
/note="assembly_fragment"

misc_feature
9605..11734
/note="assembly_fragment"

misc_feature
11835..14111
/note="assembly_fragment"

misc_feature
14212..16973
/note="assembly_fragment"

```



\* 107258 110300: contig of 3043 bp in length  
 \* 110301 110400: gap of 100 bp  
 \* 110401 113955: contig of 3555 bp in length  
 \* 113956 114055: gap of 100 bp  
 \* 114056 118801: contig of 4746 bp in length  
 \* 118802 118901: gap of 100 bp  
 \* 118902 129831: contig of 10930 bp in length  
 \* 129832 129931: gap of 100 bp  
 \* 129932 137545: contig of 7614 bp in length  
 \* 137546 137645: gap of 100 bp  
 \* 137646 146805: contig of 9160 bp in length.

## FEATURES

## source

Location/Qualifiers  
 1..146805  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="1"  
 /clone="RP11-563I16"  
 /clone.lib="RPC1-11.2"  
 1..5105  
 /note="assembly\_fragment:00368  
 clone\_end:T7  
 vector\_side:left"  
 5206..10570  
 /note="assembly\_fragment:00594  
 fragment\_chain:1"  
 10671..17626  
 /note="assembly\_fragment:00786  
 fragment\_chain:1"  
 17727..20478  
 /note="assembly\_fragment:00844  
 fragment\_chain:1"  
 20579..23035  
 /note="assembly\_fragment:00011  
 fragment\_chain:1"  
 23136..25308  
 /note="assembly\_fragment:01029  
 fragment\_chain:1"  
 25409..30469  
 /note="assembly\_fragment:00148  
 fragment\_chain:2"  
 30570..32688  
 /note="assembly\_fragment:00961  
 fragment\_chain:2"  
 32789..36187  
 /note="assembly\_fragment:00310  
 fragment\_chain:3"  
 36288..48988  
 /note="assembly\_fragment:01150  
 fragment\_chain:3"  
 49089..67333  
 /note="assembly\_fragment:00444  
 fragment\_chain:4"  
 67434..82317  
 /note="assembly\_fragment:00140  
 fragment\_chain:4"  
 82418..84614  
 /note="assembly\_fragment:00657  
 fragment\_chain:5"  
 84715..90326  
 /note="assembly\_fragment:00757  
 fragment\_chain:5"  
 90427..92847  
 /note="assembly\_fragment:00677  
 fragment\_chain:6"  
 92748..97690  
 /note="assembly\_fragment:00248  
 fragment\_chain:6"  
 97791..100799  
 /note="assembly\_fragment:00091"  
 100900..104202  
 /note="assembly\_fragment:00538"  
 104303..107157  
 /note="assembly\_fragment:00675"

misc\_feature 107258..110300  
 /note="assembly\_fragment:00695"  
 misc\_feature 110401..113955  
 /note="assembly\_fragment:01316"  
 misc\_feature 114056..118801  
 /note="assembly\_fragment:01292  
 fragment\_chain:7"  
 misc\_feature 118902..129831  
 /note="assembly\_fragment:00551  
 fragment\_chain:7"  
 misc\_feature 129932..137545  
 /note="assembly\_fragment:01410  
 fragment\_chain:7"  
 misc\_feature 137646..146805  
 /note="assembly\_fragment:00763  
 fragment\_chain:7"  
 clone\_end:SP6  
 vector\_side:right"  
 BASE COUNT 37508 a 33240 c 34076 g 39557 t 2424 others  
 ORIGIN

Query Match 26.2%; Score 349.2; DB 2; Length 146805;  
 Best Local Similarity 96.3%; Pred. No. 7.9e-65;  
 Matches 368; Conservative 0; Mismatches 13; Indels 1; Gaps 1;  
 QY 709 GACAGAGACTCTGCTCAAAAAGGACATTTATCATATATACATCTTATTAGAGCCCC 768  
 |||||  
 Db 84740 GACAGAGACTCTGCTCAAAAAGGACATTTATCATATATACATCTTATTAGAGCCCC 84799  
 QY 769 TAATTTCTATCTGAAGGCACCTGTTTTTTTTTTTAAACAGTTAAGTACTGATCTCAACAG 828  
 |||||  
 Db 84800 TAATTTCTATCTGAAGGCACCTGTTTTTTTTTTTAAACAGTTAAGTACTGATCTCAACAG 84859  
 QY 829 ACAATATTTCTGATCAGATAGTCCCTCTCAACAGTAGCAAAATGTTGTTTCATAAAGTG 888  
 |||||  
 Db 84860 ACAATATTTCTGATCAGATAGTCCCTCTCAACAGTAGCAAAATGTTGTTTCATAAAGTG 84919  
 QY 889 GGAAGAAACAGCATTTTAAAGTAACTTTTGGGAGACTGATTTGAGTAAATAAAACT 948  
 |||||  
 Db 84920 GGAAGAAACAGCATTTTAAAGTAACTTTTGGGAGACTGATTTGAGTAAATAAAACT 84978  
 QY 949 CTGGTCTCCCTTAAGAAAAAACCCCTTCCACCTTTACTGTCTATTTATATATCCCTTTA 1008  
 |||||  
 Db 84979 CTGGTCTCCCTTAAGAAAAAACCCCTTCCACCTTTACTGTCTATTTATATATCCCTTTA 85038  
 QY 1009 GTTCCAAAGTTAATTTATCTTATTTCTGGATATTTGCTTTTATACCAAGAGCCCTTATCAGC 1068  
 |||||  
 Db 85039 GTTCCAAAGTTAATTTATCTTATTTCTGGATATTTGCTTTTATACCAAGAGCCCTTATCAGC 85098  
 QY 1069 CAGTTCCAGAACCAACCACTATA 1090  
 |||||  
 Db 85099 CCTTGTAACTACAGTATCTTTA 85120

RESULT 5  
 HUMSTTRNA  
 LOCUS Human statherin mRNA, complete cds.  
 DEFINITION Human statherin mRNA, complete cds.  
 ACCESSION M18371  
 VERSION M18371.1 GI:338610  
 KEYWORDS statherin.  
 SOURCE Human female submandibular gland, cDNA to mRNA, clone pBRHSMF988.2.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 542)  
 Dickinson, D.P., Ridall, A.L. and Levine, M.J.  
 Human submandibular gland statherin and basic histidine-rich  
 peptide are encoded by highly abundant mRNA's derived from a common  
 ancestral sequence  
 JOURNAL Biochem. Biophys. Res. Commun. 149 (2), 784-790 (1987)  
 MEDLINE 88106506  
 PUBMED 3426601



Wagatsuma, M., Murakawa, K., Kanehori, K., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahari, K., Masuho, Y., Nagai, K. and Isogai, T.  
 NEDO human cDNA sequencing project  
 Unpublished  
 2 (bases 1 to 1584)  
 Isogai, T. and Yamamoto, J.  
 Direct Submission  
 Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan  
 (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)  
 NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.

## FEATURES

Location/Qualifiers  
 1. 1584  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="SALGL1000107"  
 /tissue\_type="salivary gland"  
 /clone\_lib="SALGL1"  
 /notes="cloning vector: pME18SFL3"  
 89. .997  
 /notes="unnamed protein product"  
 /codon\_start=1  
 /protein\_id="BAC03943.1"  
 /db\_xref="GI:21751328"  
 /translation="MGDDSEMLKLPVDQKEHLKWLKARLSGYEALKIFQKINDEKSP  
 EWSKGLIKFTYIDSNVQLKLEALVYVENAHVAGTKTGEVTVGVSVFNOPK  
 AKAEGLIEICLMYIEIEGEAVQELLKGNPKIIVACITLTKALSRFGSKII  
 LKPIIKVLPKLESRAVDEAKLIAVEIYRWIDALRPPLONINSVOLKELEEM  
 VKLPTSAFPTRELRSQEALKELEQSQSGDAGEGGDDGDEVPDIDAYELLEAVEI  
 LSKLPKTFMKLQKNGKREKRWSL"

BASE COUNT 526 a 298 c 331 g 429 t  
 ORIGIN

Query Match 19.0%; Score 253; DB 9; Length 1584;  
 Best Local Similarity 88.4%; Pred. No. 6.6e-44;  
 Matches 274; Conservative 0; Mismatches 36; Indels 0; Gaps 0;  
 QY 1022 TTATCTTATTCTGGATATGCTTTTATACCAAGAGCCCTTATCAGCCAGTTCAGAAC 1081  
 Db 1174 TTTGCGTAGAATTGGAAGATTCGTTATGGGTATGGCCCTTATCAGCCAGTTCAGAAC 1233  
 QY 1082 ACCACTATACGCAACCAACCAACCAATACCAATACATATACGTTTAAATATCATCA 1141  
 Db 1234 ACCACTATACCAACCAACCAACCAATACCAATACATATACGTTTAAATATCATCA 1293  
 QY 1142 GTAACCTGAGGACATGATTATTGAGGCTTGATTGGCAATACGACTTCTACATCCATTT 1201  
 Db 1294 GTAACCTGAGGACATGATTATTGAGGCTTGATTGGCAATACGACTTCTACATCCATTT 1353  
 QY 1202 CTATCTTTTATACCATATACATACATACATCTTTTGTGAGATCATCTPAAGAGCAATG 1261  
 Db 1354 CTATCTTTTATACCATATACATACATACATCTTTTGTGAGATCATCTPAAGAGCAATG 1413  
 QY 1262 CGAATGTAACCAACCTTAATTTTACTGGATCTCTTTGGTCCAGATCTTGGCTTTTCCA 1321  
 Db 1414 CAATGTAACCAACCTTAATTTTACTGGATCTCTTTGGTCCAGATCTTGGCTTTTCCA 1473  
 QY 1322 ATGTCACCTTG 1331  
 Db 1474 TTGTCACCTTG 1483

RESULT 8  
 G05711  
 LOCUS

G05711 297 bp DNA 11linear STS 19-OCT-1995

## DEFINITION

human STS WI-7844, sequence tagged site.

G05711

VERSION

ST5: STS sequence; primer; sequence tagged site.

KEYWORDS

SOURCE

collection.

## ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 297)

AUTHORS

Whitehead Institute/MIT Center for Genome Research; Physically

Mapped ESTs

Unpublished (1995)

JOURNAL

COMMENT

Contact: Thomas Hudson

Whitehead Institute/MIT Center for Genome Research

Whitehead Institute for Biomedical Research

9 Cambridge Center, Cambridge MA 02142 USA

Tel: 617 252 1900

Fax: 617 252 1902

Email: thudson@genome.wi.mit.edu

Primer A: ACTGACGACATGATTTATTGAGG

Primer B: AAAAATGTGCTTGAATTTATTGC

STS size: 286

PCR profile:

Presoak:

Denaturation:

Annealing: 56 degrees C

Polymerization:

PCR Cycles: 35

Thermal Cycler:

Protocol:

Template: 10 ng

Primer: each 5 pm

dNTPs: each 4 mM

Taq Polymerase: 0.025 units/ul

Total Vol: 20 ul

Buffer:

MgCl2: 1.5 mM

KCl: 50 mM

Tris-HCl: 10 mM

pH: 9.3

Prepared with primer pairs derived from M18371 -- Unigene.

Location/Qualifiers

1. .297

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/map="748\_F.3; 793\_E.2; 794\_A.9; 886\_B.3; 960\_D.2;

972\_E.4; 750\_D.11; 855\_E.12; 882\_G.11; 931\_G.10; 395.8 cr

from top of Chr4 linkage group"

12. .297

primer\_bind 12. .34

BASE COUNT 99 a 56 c 37 g 105 t

ORIGIN

Query Match 13.2%; Score 176.2; DB 11; Length 297;

Best Local Similarity 92.9%; Pred. No. 2.2e-27;

Matches 184; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 1134 TATCATCACTAATCGAGGACATGATTATTGAGGCTTGATGGCAATACGACTTCTACA 1193

Db 1 TATCATCACTAATCGAGGACATGATTATTGAGGCTTGATGGCAATACGACTTCTACA 60

QY 1194 TCCATATTTCTATCTTTTATACCATATACATACATACATACATACATACATACAT 1253

Db 61 TCCATATTTCTATCTTTTATACCATATACATACATACATACATACATACATACAT 120

QY 1254 GAGCAATGCAATGTAAACCCCTATAATTTACTGGATCTCTTTGGTCCAGATCTTC 1313

```

||||| 121 GAGCAATGCAATGAAACACTATAATTTACTGTATACCTTTGTTTCAGGATCTGC 180
||||| 1314 CTTTCCCAATGTCACCTTG 1331
||||| 181 CTTTCCCAATGTCACCTTG 198

RESULT 9
LOCUS      G26643              1498 bp      DNA      linear      STS 02-JUN-1996
DEFINITION human STS_M32639, sequence tagged site.
ACCESSION  G26643
VERSION    G26643.1 GI:1348875
KEYWORDS   STS; STS sequence; primer; sequence tagged site.
SOURCE     Homo sapiens STSs derived from sequences in dbEST and the Unigene
           collection.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 1498)
AUTHORS   Hudson,T.
TITLE     Whitehead Institute/MIT Center for Genome Research; Physically
JOURNAL   Mapped STS
COMMENT   Unpublished (1995)

Contact: Thomas Hudson
Whitehead Institute/MIT Center for Genome Research
Whitehead Institute for Biomedical Research
9 Cambridge Center, Cambridge MA 02142 USA
Tel: 617 252 1900
Fax: 617 252 1902
Email: thudson@genome.wi.mit.edu

Primer A: GGACATGATTATTCAGCTAAGATGG
Primer B: ACTGCCTCTACCATAGCAGG
STS size: 210
PCR Profile:
  Presoak:
  Denaturation:
  Annealing: 56 degrees C
  Polymerization:
  PCR Cycles: 35
  Thermal Cycler:
Protocol:
  Template: 10 ng
  Primer: each 5 pM
  dNTPs: each 4 nM
  Taq Polymerase: 0.025 units/ul
  Total Vol: 20 ul

Buffer:
  MgCl2: 1.5 mM
  KCl: 50 mM
  Tris-HCl: 10 mM
  pH: 9.3

Derived from dbEST (genbank accession M32639).
FEATURES             Location/Qualifiers
     source           1..1498
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="473.8 CR from top of Chr4 linkage group"
     STS              21..230
     primer_bind      21..45
     primer_bind      complement(209..230)
     BASE COUNT       511 a 229 c 244 g 514 t
     ORIGIN
Query Match          10.98; Score 145.4; DB 11; Length 1498;
Best Local Similarity 90.19; Pred. No. 6.1e-21;
Matches 155; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

```

```

QY 1160 TATTGAGCTTGATTGGCAATACGACTTCTACATCCATATTCATCTTCATACCATA 1219
Db 1234 TGTGAGGCTTGATTGGCAATACGACTTCTACATCCATATTCATCTTCATACCATA 1293
QY 1220 TCACACTACTACCACCTTTTGTNAGATCATCTAAGAGCAATGCAATGTAAACCCCTATA 1279
Db 1294 TCACACTACTACCACCTTTTGAAGAATCATCAAGAGCAATGCAATGAANAACACTATA 1353
QY 1280 ATTTACTGGATACTCTTTGGTCCAGATACTTGCCTTTTCCAATGTCACTTG 1331
Db 1354 ATTTACTGTATACTCTTTGTTTCAGGATACTTGCCTTTTCAATTTGTCACTTG 1405

RESULT 10
LOCUS      HUMSTATH2              4723 bp      DNA      linear      PRI 13-JAN-1995
DEFINITION Human salivary statherin gene, exons 2-6.
ACCESSION  M32639
VERSION    M32639.1 GI:338504
KEYWORDS   statherin.
SEGMENT    2 of 2
SOURCE     Human (individuals #563, #8136, and J.F.) fibroblast, cell line
           #563, DNA, clones 1-3.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 4723)
AUTHORS   Sabatini,L.M., He,Y.Z. and Azen,E.A.
TITLE     Structure and sequence determination of the gene encoding human
           salivary statherin
JOURNAL   Gene 89 (2), 245-251 (1990)
MEDLINE   90323623
PubMed    2373369
COMMENT    Draft entry and computer-readable sequence for [1] kindly submitted
           by L.M.Sabatini, 03-JAN-1990.
FEATURES             Location/Qualifiers
     source           1..4723
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="4q11-q13"
                     join(M31077.1:1601..1931,1..352)
                     /gene="STATH"
     prim_transcript  <1..4309
                     /note="STATH mRNA and introns"
                     <1..286
                     /gene="STATH"
                     /note="STATH intron A"
                     join(302..352,1532..1552,1645..1674,2728..2814)
                     /note="statherin precursor"
                     /codon_start=1
                     /protein_id="AAA60593.1"
                     /db_xref="GI:338506"
                     /translation="MKFLVFATLALMVSMIGADSSSEKFLRRTRGFGYGVQYPVP
                     EOPLYPOPYQOYQOYTF"
     sig_peptide      join(302..352,1532..1537)
                     /note="statherin signal peptide"
     mat_peptide      join(1538..1552,1645..1674,2728..2811)
                     /product="statherin"
                     <302..352
                     /gene="STATH"
                     /note="statherin precursor, (first translated exon);
                     G00-120-391"
                     /number=2
     intron           353..1531
                     /note="STATH intron B"
     exon             1532..1552
                     /number=3
     intron           1553..1644
                     /note="STATH intron C"
     exon             1645..1674
                     /number=4
     intron           1675..2727
                     /note="STATH intron D"

```











Search completed: April 14, 2003, 22:34:32  
Job time : 4275 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 14, 2003, 19:59:10 ; Search time 330 Seconds  
(without alignments)  
9083.060 Million cell updates/sec

Title: US-09-830-244B-2

Perfect score: 1331

Sequence: 1 catgttttagaatcaag.....gcctttccaatgtcacttg 1331

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_101002.\*  
1: /SID22/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SID22/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SID22/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SID22/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SID22/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SID22/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SID22/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SID22/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SID22/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SID22/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SID22/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SID22/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SID22/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SID22/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SID22/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SID22/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SID22/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SID22/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SID22/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SID22/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SID22/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SID22/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1330	99.9	1331	21	AAA48963 Human lysine-rich
2	253	19.0	552	21	AAA48964 Human statherin DN
3	209.6	15.7	2121	21	AAC76665 Human ORFX ORF2220
4	187.4	14.1	221	21	AAA42335 Human secreted exp
5	86.6	6.5	390	22	AA191635 Human polynucleoti
6	86.2	6.5	857	22	AAH98658 Human EST-derived
7	78.8	5.9	491	21	AAA48965 Human basic histid
8	78.6	5.9	126	16	AAW24320 Human gene signatu
9	76.4	5.7	90	16	AAW23140 Human gene signatu

10	74.6	5.6	9881	24	ABL54353	Chemically treated
11	60.2	4.5	516	20	AAV89525	EST clone Cp289.
12	56.4	4.2	60	24	ABN43273	Human spliced tran
13	55.8	4.2	9881	24	ABL54354	Chemically treated
14	48.4	3.6	6056	24	ABL33026	Human immune syste
15	47	3.5	61020	22	AA346787	Tumour suppressor
16	46.8	3.5	10279	24	ABL92276	Chemically treated
17	46.8	3.5	10279	24	ABL33590	Human immune syste
18	46.8	3.5	10279	24	AA22327	Human immune syste
19	46.8	3.5	513445	22	AA161373	Chemically treated
20	46	3.5	6131	24	ABL31373	Soybean 318013 reg
21	44.8	3.4	545	22	ABL32891	Human immune syste
22	44.6	3.4	556	22	AA191773	Human cervical can
23	44	3.3	16766	24	ABL34157	Human polynucleoti
24	43.8	3.3	5461	24	ABQ66977	Human immune syste
25	43.8	3.3	5461	24	ABL32998	Human anglogenesis
26	43.8	3.3	5647	24	ABL70355	Human immune syste
27	43.8	3.3	5647	24	ABL33566	Chemically treated
28	43.8	3.3	5647	24	ABL33566	Human immune syste
29	43.4	3.3	8577	24	ABL33787	Human gene regulat
30	43.4	3.3	19659	24	ABL32766	Human immune syste
31	42.8	3.2	375	23	ABV44911	Human immune syste
32	42.8	3.2	8946	24	ABL32910	Human prostate exp
33	42.6	3.2	3680	22	AAH14565	Human immune syste
34	42.6	3.2	7823	22	AAH72614	Human CDNA sequenc
35	42.4	3.2	18183	22	AA546279	Human cervical can
36	42.4	3.2	18183	24	ABL70111	Tumour suppressor
37	42.4	3.2	18183	24	ABK31158	Chemically treated
38	42.2	3.2	1160	19	AAV44863	Signal transductio
39	42.2	3.2	1160	22	AAF98460	Clone CK201-1 codi
40	42.2	3.2	6055	22	AA36246	Human CDNA clone C
41	42.2	3.2	7809	22	AA36244	Human cardiovascu
42	42.2	3.2	7809	22	AA36244	Human cardiovascu
43	42	3.2	707	22	AA24056	Human breast cance
44	42	3.2	15373	24	ABL32466	Human immune syste
45	41.8	3.1	7171	24	ABN80216	Human chemically m

#### ALIGNMENTS

RESULT 1  
AAA48963  
ID AAA48963 standard; CDNA; 1331 BP.

AC AAA48963;

XX 06-OCT-2000 (first entry)

XX Human lysine-rich statherin CDNA from Incyte clone 2820214.

XX Lysine-rich statherin protein; LRSP; acidic phosphoprotein; human;  
XX precipitation inhibitor; autoimmunity; inflammatory disorder; AIDS;  
XX asthma; allergy; diabetes mellitus; fungal; bacterial infection;  
XX cancer; leukemia; adenocarcinoma; melanoma; ss.

OS Homo sapiens.

XX Key Location/Qualifiers  
XX CDS 871..1158

XX FT /\*tag= a  
XX FT /product= Lysine-rich\_statherin

XX PN WO200024779-A1.

XX XX 04-MAY-2000.

XX XX 22-OCT-1999; 99WO-US24046.

XX XX 23-OCT-1998; 98US-0155209.

XX XX (INCY-) INCYTE PHARM INC.

PI Tang YT, Corley NC, Guegler KJ, Patterson C;  
 XX WPI: 2000-350699/30.  
 DR P-PSDB; AAY94526.  
 XX  
 PT Purified polypeptide used for treating or preventing a disorder  
 PT characterized by expression or activity of lysine-rich statherin  
 PT proteins -  
 PS  
 PS Claim 3; Page 70; 75pp; English.  
 XX  
 CC The present sequence is human lysine-rich statherin protein (LRSP)  
 CC cDNA from Incyte clone 2820214. This sequence was identified  
 CC through analysis of a cDNA library of breast tumour tissue  
 CC (BRSTNOR14). The LRSP sequence was found to have homology with  
 CC human statherin (AAY94527) and human basic histidine-rich protein  
 CC (AAY94528). Human statherin is a phosphoprotein that acts as an  
 CC inhibitor of precipitation of calcium phosphate salts in the oral  
 CC cavity. The LRSP polypeptide and its antagonists may be useful for  
 CC treating or preventing disorders associated with the activity of  
 CC LRSP. Such disorders include autoimmune/inflammatory disorders (for  
 CC example AIDS, allergies, asthma, diabetes mellitus), bacterial and  
 CC fungal infection and cancers (such as leukemia, adenocarcinoma,  
 CC melanoma). Antibodies to LRSP may be useful for diagnosis of  
 CC the above disorders.  
 XX  
 SQ Sequence 1331 BP; 397 A; 258 C; 228 G; 447 T; 1 other;  
 Query Match 99.9%; Score 1330; DB 21; Length 1331;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTATGTTTTAGAAATCAAGATGAACGGTAAAGCTGCTCATGTACCAACGTTGAATTT 60  
 DB 1 CTATGTTTTAGAAATCAAGATGAACGGTAAAGCTGCTCATGTACCAACGTTGAATTT 60

QY 61 ACAGTGTTCACAAATGCTGGAATTTTGCACCTGCCATAGGGAATGTTAAGTTACTTGGC 120  
 DB 61 ACAGTGTTCACAAATGCTGGAATTTTGCACCTGCCATAGGGAATGTTAAGTTACTTGGC 120

QY 121 TGGAAATTTACAGCTTGTGAGTAAACAAGTTGAAGTTAGCAGATGAGGGGAATATTG 180  
 DB 121 TGGAAATTTACAGCTTGTGAGTAAACAAGTTGAAGTTAGCAGATGAGGGGAATATTG 180

QY 181 AGSCCCTAAGGCTTAACAAATAATCAGTATCTGAGATAGTGGCTAATGTGGCTCCCA 240  
 DB 181 AGSCCCTAAGGCTTAACAAATAATCAGTATCTGAGATAGTGGCTAATGTGGCTCCCA 240

QY 241 GGCCTAATTTGGGAACAGTTTTCCTGATTGCTTTGAGAAGTACTTCTTTTGACAGAAA 300  
 DB 241 GGCCTAATTTGGGAACAGTTTTCCTGATTGCTTTGAGAAGTACTTCTTTTGACAGAAA 300

QY 301 TTTTCATTCGCTTGCCATTGCTAATTCCTCTTTATAGGAGCCATTGATTTCTTCC 360  
 DB 301 TTTTCATTCGCTTGCCATTGCTAATTCCTCTTTATAGGAGCCATTGATTTCTTCC 360

QY 361 TTTTGTGGGAATGTCCTCAATAGCATTTTCAGATCTTTTGTGATGTCACCTAATGCCATTAT 420  
 DB 361 TTTTGTGGGAATGTCCTCAATAGCATTTTCAGATCTTTTGTGATGTCACCTAATGCCATTAT 420

QY 421 TGGTAATGCGGTATTGGTGAATACAGCATAGTAAATTAACCTGTTACAGTAATCTACA 480  
 DB 421 TGGTAATGCGGTATTGGTGAATACAGCATAGTAAATTAACCTGTTACAGTAATCTACA 480

QY 481 CTGGATTTGCTGCACCTCTACCAATAGCCTTTTGAATGACCTGAAGTCTTAACAGAGAA 540  
 DB 481 CTGGATTTGCTGCACCTCTACCAATAGCCTTTTGAATGACCTGAAGTCTTAACAGAGAA 540

QY 541 AGAGGATCTCTGCAGAAAGATAGCTAATATTTTGGTACTTTATCTGAAATCCAG 600  
 DB 541 AGAGGATCTCTGCAGAAAGATAGCTAATATTTTGGTACTTTATCTGAAATCCAG 600

QY 601 ATGCTGCTTCCCTGAGGTGTTTTCCTTCTTACGATCCTCATGAAATCCCTCTCGGA 660

Db 601 ATGCTGCTTCCCTGAGGTGTTTTCCTTCTTACGATCCTCATGAAATCCCTCTCGGA 660  
 QY 661 GCACAGGACAGTATAGTAACTCTCCATTTCTTTGTTTGTGTTTAAAGACAGACTCT 720  
 Db 661 GCACAGGACAGTATAGTAACTCTCCATTTCTTTGTTTGTGTTTAAAGACAGACTCT 720  
 QY 721 GTCTCAAAAAAAGGACATTTATCATATTAACATCTTATTAGAGCCCTTAATTTCTATC 780  
 Db 721 GTCTCAAAAAAAGGACATTTATCATATTAACATCTTATTAGAGCCCTTAATTTCTATC 780  
 QY 781 TGAAGGCACTGTTTATTTTAAACAGTTAACTAGTACTGATGTCACAGACAAATATTTCT 840  
 Db 781 TGAAGGCACTGTTTATTTTAAACAGTTAACTAGTACTGATGTCACAGACAAATATTTCT 840  
 QY 841 GATCAGATAGTCCCTGTCACAGTAGCAAAATGCTGTTTCATATAAGTGGGAAGAAACAG 900  
 Db 841 GATCAGATAGTCCCTGTCACAGTAGCAAAATGCTGTTTCATATAAGTGGGAAGAAACAG 900  
 QY 901 CATTTTAAAGTAACTTTTGGGAGACTGATTTGAGTAATAATAAACTCTGGTCTCCCTT 960  
 Db 901 CATTTTAAAGTAACTTTTGGGAGACTGATTTGAGTAATAATAAACTCTGGTCTCCCTT 960  
 QY 961 AAGAAAAAACCCTTCCACCTTTACTGTGTCATTTATATCCCTTAGTTCACAAAGTTA 1020  
 Db 961 AAGAAAAAACCCTTCCACCTTTACTGTGTCATTTATATCCCTTAGTTCACAAAGTTA 1020  
 QY 1021 ATTATCTATTCTGATATTTGCTTTTATACCAAGAGCTTTATCAGCCAGTTCACAGAAC 1080  
 Db 1021 ATTATCTATTCTGATATTTGCTTTTATACCAAGAGCTTTATCAGCCAGTTCACAGAAC 1080  
 QY 1081 AACCACTATACGCAACCAATACCAACCAATACCAACCAATATACGTTTAAATATCATC 1140  
 Db 1081 AACCACTATACGCAACCAATACCAACCAATACCAACCAATATACGTTTAAATATCATC 1140  
 QY 1141 AGTAAGTGCAGGACATGATTTAGTGGCTTGTGTTGCAAAATACGACTTCTACATCCATAT 1200  
 Db 1141 AGTAAGTGCAGGACATGATTTAGTGGCTTGTGTTGCAAAATACGACTTCTACATCCATAT 1200  
 QY 1201 TCTCATCTTTCATACCATATACACTACTACCATCTTTTGTNAGATCATCTAAGAGCAAT 1260  
 Db 1201 TCTCATCTTTCATACCATATACACTACTACCATCTTTTGTNAGATCATCTAAGAGCAAT 1260  
 QY 1261 GCGAATGTAAACCCCTATAATTTACTGGTACTCTTTGGTTCAGATACTTGCCTTTCC 1320  
 Db 1261 GCGAATGTAAACCCCTATAATTTACTGGTACTCTTTGGTTCAGATACTTGCCTTTCC 1320  
 QY 1321 AATGTCACCTTG 1331  
 Db 1321 AATGTCACCTTG 1331

RESULT 2  
 AAA48964  
 ID AAA48964 standard; DNA; 552 BP.  
 XX  
 AC AAA48964;  
 XX  
 DT 06-OCT-2000 (first entry)  
 XX  
 DE Human statherin DNA.  
 XX  
 KW Lysine-rich statherin protein; LRSP; acidic phosphoprotein; human;  
 KW precipitation inhibitor; autoimmune; inflammatory disorder; AIDS;  
 KW asthma; allergy; diabetes mellitus; fungal; bacterial infection;  
 KW cancer; leukemia; adenocarcinoma; melanoma; ds.  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 73..261  
 FT /tag= a  
 FT /product= Statherin

XX WO200024779-A1.  
 XX 04-MAY-2000.  
 XX 22-OCT-1999; 99WO-US24046.  
 XX 23-OCT-1998; 98US-0155209.  
 XX (INCY-) INCYTE PHARM INC.  
 XX Tang YT, Corley NC, Guegler KJ, Patterson C;  
 XX WPI; 2000-350699/30.  
 XX P-PSDB; AAY94527.  
 XX Purified polypeptide used for treating or preventing a disorder  
 XX characterized by expression or activity of lysine-rich statherin  
 XX proteins -  
 XX  
 XX Disclosure; Page 71; 75pp; English.  
 XX  
 XX The present invention relates to human lysine-rich statherin protein  
 XX (LRSP)(AAY94526). The cDNA sequence encoding this protein was identified  
 XX through analysis of a cDNA library of breast tumour tissue (BRSTNOT14).  
 XX The LRSP sequence was found to have homology with the DNA of human  
 XX statherin (the present sequence) and human basic histidine-rich protein  
 XX (AAY94528). Human statherin is a phosphoprotein that acts as an  
 XX inhibitor of precipitation of calcium phosphate salts in the oral cavity.  
 XX The LRSP polypeptide and its antagonists may be useful for treating or  
 XX preventing disorders associated with the activity of LRSP. Such  
 XX disorders include autoimmune/inflammatory disorders (for example AIDS,  
 XX allergies, asthma, diabetes mellitus), bacterial and fungal infection  
 XX and cancers (such as leukemia, adenocarcinoma, melanoma). Antibodies to  
 XX LRSP may be useful for diagnosis of the above disorders.  
 XX  
 XX Sequence 552 BP; 172 A; 122 C; 78 G; 180 T; 0 other;

Query Match 19.0%; Score 253; DB 21; Length 552;  
 Best Local Similarity 88.4%; Pred. No. 2e-53;  
 Matches 274; Conservative 0; Mismatches 36; Indels 0; Gaps 0;  
 QY 1022 TTATCTTATTCGGATATGCTTTTATACCAAGAGCCTTATCAGCCAGTTCAGAACCA 1081  
 DB 150 TTTCGTAGAAITGGGAAGATTCGGTTATGGGTATGGCCTTATCAGCCAGTTCAGAACCA 209  
 QY 1082 ACCACTATACGCACACCAATACCAACCAACCAATACCAATACCAATACCAATACCAAT 1141  
 DB 210 ACCACTATACGCACACCAATACCAACCAACCAATACCAATACCAATACCAATACCA 269  
 QY 1142 GTAACTGCAGGACATGATTATTGAGGCTTGATTGGCAATACGACTTCTACATCCATATT 1201  
 DB 270 GTAACTGCAGGACATGATTATTGAGGCTTGATTGGCAATACGACTTCTACATCCATATT 329  
 QY 1202 CTCATCTTTTCATACCATATCAGACTACTACCATCTTTTGTGAGATCATCATAGAGCAATG 1261  
 DB 330 CTCATCTTTTCATACCATATCAGACTACTACCATCTTTTGTGAGATCATCATAGAGCAATG 389  
 QY 1262 CGAATGAAACCCCTATATTTACTGATGACTCTTTGTTCCAGATACCTTGCTTTTCA 1321  
 DB 390 CAATGAAACCACTATATTTACTGATGACTCTTTGTTCCAGATACCTTGCTTTTCA 449  
 QY 1322 ATGTCACCTTG 1331  
 DB 450 TTGTCACCTTG 459  
 RESULT 3  
 AAC76665  
 ID AAC76665 standard; CDNA; 2121 BP.  
 XX  
 AC AAC76665;  
 XX

DT 08-FEB-2001 (first entry)  
 XX Human ORFX ORF2220 polynucleotide sequence SEQ ID NO:4439.  
 DE  
 XX  
 XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
 XX vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 KW hypotensive; dermatological; immunosuppressive; antineoplastic;  
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;  
 KW neurodegenerative disease; osteoarthritis; graft vs host disease;  
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KW cholesterol ester storage; systemic lupus erythematosus; infection;  
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
 KW thrombosis; contraceptive; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200058473-A2.  
 XX  
 XX 05-OCT-2000.  
 XX  
 XX 31-MAR-2000; 2000WO-US08621.  
 XX  
 XX 31-MAR-1999; 99US-0127607.  
 XX 02-APR-1999; 99US-0127636.  
 XX 05-APR-1999; 99US-0127728.  
 XX 30-MAR-2000; 2000US-0540763.  
 XX (CURA-) CURAGEN CORP.  
 XX  
 XX Shinkets RA, Leach M;  
 DR WPI; 2000-602362/57.  
 DR P-PSDB; AAB42456.  
 XX  
 XX Novel nucleic acids and peptides derived from open reading frame X,  
 XX useful for treating e.g. cancers, proliferative disorders,  
 XX neurodegenerative disorders and cardiovascular disease -  
 XX  
 XX Claim 5; Page 3631-3632; 5507pp; English.  
 XX  
 XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
 XX which represent the human ORFX open reading frames 1 to 3161. The ORFX  
 XX sequences have activities such as: cytostatic; hepatotropic; vulnary;  
 XX antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
 XX osteopathic; anticonvulsant; antiarthritic; coagulant; immunosuppressant;  
 XX immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
 XX antidiabetic; hypotensive; dermatological; immunosuppressive;  
 XX antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;  
 XX antithyroid; and antianaemic. The sequences can be used for determining  
 XX the presence of or predisposition to, or preventing or treating  
 XX pathological conditions associated with an ORFX-associated disorder. The  
 XX nucleic acids can be used to express ORFX proteins in gene therapy  
 XX vectors. The proteins and nucleic acids may be used to treat cancers,  
 XX proliferative disorders, neurodegenerative disorders, osteoarthritis,  
 XX graft vs host disease, cardiovascular disease, diabetes mellitus,  
 XX hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
 XX erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
 XX bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
 XX allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
 XX nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
 XX coagulation; to inhibit thrombosis; and as a contraceptive.  
 XX  
 XX Sequence 2121 BP; 617 A; 430 C; 341 G; 732 T; 1 other;

Query Match 15.7%; Score 209.6; DB 21; Length 2121;  
 Best Local Similarity 81.5%; Pred. No. 2.2e-42;  
 Matches 255; Conservative 0; Mismatches 55; Indels 3; Gaps 1;



QY 1022 TTATCTTATTTTGGATATGCTTTTATACCAAGAGCCCTTATCAGCCAGTTCCAGAAC 1081  
 Db 1578 TTGGCGTAGAATGGGAAGATTCGGTTATGGGTATGCGCCCTTATCAGCCAGTTCCAGAAC 1637  
 QY 1082 ACCACTATACGCACACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 1141  
 Db 1638 ACCACTATACGCACACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 1697  
 QY 1142 GTAACGAGGACATGATTATGAGGCTTGTGAGTGGCAATACGACCTTCTACATCCATAT 1201  
 Db 1698 GTAACGAGGACATGATTATGAGGCTTGTGAGTGGCAATACGACCTTCTACATCCATAT 1757  
 QY 1202 CTCATCTTTTCATACCATACACATACACATACACATACACATACACATACACATAC 1261  
 Db 1758 CTCATCTTTTCATACCATACACATACACATACACATACACATACACATACACATAC 1817  
 QY 1262 CGAATGTA---AAACCCCTATATTTACTGGGATACCTTTTGGTCCAGATACCTTGGCTTT 1318  
 Db 1818 CAGATAAAGAAATACCATGATTACTGATACCTTTTGGTCCAGATACCTTGGCTTT 1877  
 QY 1319 CCAATGTCACATG 1331  
 Db 1878 TAATTATCATTTG 1890

## RESULT 4

AA42355  
 ID AA42355 standard; cDNA; 221 BP.

XX  
 AC AA42355;

XX  
 DT 21-AUG-2000 (first entry)

XX  
 DE Human secreted expressed sequence tag SEQ ID NO:1095.

XX  
 KW Human; mouse; xenopus; rat; secreted expressed sequence tag; SEST;  
 KW expressed sequence tag; EST; probe; chemotactic; proliferative;  
 KW immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;  
 KW thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;  
 KW antiviral; antidiabetic; antiasthmatic; vulnary; antiparkinsonian;  
 KW antitumor; osteopathic; neuroprotective; neurotropic; antiproliferative;  
 KW cerebroprotective; anticonvulsant; antidepressant; gene therapy;  
 KW vaccine; autoimmune diabetes; asthma; myeloid cell deficiency; ulcer;  
 KW insulin dependent diabetes; multiple sclerosis; allergic condition;  
 KW lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;  
 KW central nervous system disorder; Alzheimer's disease; stroke;  
 KW Parkinson's disease; Huntington's disease; coagulation disorder;  
 KW haemophilia; thrombosis; inflammatory disorder; Crohn's disease;  
 KW tumour; infection; depression; psoriasis; ss.

XX  
 OS Homo sapiens.

XX  
 PN WO200021990-A1.

XX  
 PD 20-APR-2000.

XX  
 PF 15-OCT-1999; 99WO-US24205.

XX  
 PR 15-OCT-1998; 98US-0104435.

XX  
 PA (GEM) GENETICS INST INC.

XX  
 PI Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;  
 PI Merberg D, Treacy M;

DR  
 WPI; 2000-317937/27.

XX  
 PT Isolated polynucleotides, and encoded proteins, comprising secreted  
 PT expressed sequence tags (SESTs), useful for treating various disorders  
 PT such as autoimmune, infectious, and central nervous system disorders -

XX  
 PS Claim 1; Page 394; 618pp; English.

XX

CC AA411261 to AA43419 represent specifically claimed secreted expressed  
 CC sequence tags (SESTs), isolated from human, mouse, xenopus and rat  
 CC tissue sources. The SESTs can have a range of activities depending on  
 CC the tissues they were isolated from. The activities include:  
 CC chemotactic; proliferative; immunomodulatory; haematopoietic;  
 CC chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory;  
 CC cytostatic; antibacterial; antifungal; antitumor; antidiabetic;  
 CC antiasthmatic; antidiabetic; antitumor; osteopathic; neuroprotective;  
 CC neurotropic; antiparkinsonian; antiproliferative; cerebroprotective;  
 CC anticonvulsant; and antidepressant. The SESTs can be used for gene  
 CC therapy and in vaccines. The SESTs are useful as probes for the  
 CC identification and isolation of full-length cDNAs and genomic DNA  
 CC molecules which correspond to the SESTs. Proteins encoded by the SESTs  
 CC are useful in assays for determining biological activity and raising  
 CC antibodies. They may be useful for treatment of autoimmune disorders  
 CC (multiple sclerosis, insulin dependent diabetes), allergic conditions  
 CC (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,  
 CC osteoporosis, osteoarthritis, central nervous system disorders  
 CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation  
 CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's  
 CC disease), tumours, bacterial, fungal or viral infections, depression and  
 CC psoriasis. AA43420 to AA43425 represent linker variants which are given  
 CC in the exemplification of the present invention.

XX  
 SQ Sequence 221 BP; 72 A; 41 C; 44 G; 64 T; 0 other;

Query Match 14.1%; Score 187.4; DB 21; Length 221;  
 Best Local Similarity 99.5%; Pred. No. 3.6e-37;  
 Matches 188; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 793 TTTTNTTTTAAACAGTTAAGTACTGATGTCACACAGACAAATATTTCTGATCAGATGTC 852  
 Db 28 TTTTNTTTTAAACAGTTAAGTACTGATGTCACACAGACAAATATTTCTGATCAGATGTC 87

QY 853 CCCTGTCAACAGTAGCAAAATGTGGTTTCATAAAGTGGAGAAACAGACATTTTAAAGTA 912

Db 88 CCCTGTCAACAGTAGCAAAATGTGGTTTCATAAAGTGGAGAAACAGACATTTTAAAGTA 147

QY 913 ACTTTTGGGAGACTGATTTGAGTAATAATAAACTCTGGTCTCCCTTAAGAAAAA 972

Db 148 ACTTTTGGGAGACTGATTTGAGTAATAATAAACTCTGGTCTCCCTTAAGAAAAA 207

QY 973 CCCTTCCAC 981

Db 208 CCCTTCCGC 216

## RESULT 5

AAI91635

ID AAI91635 standard; cDNA; 390 BP.

XX  
 AC AAI91635;

XX  
 DT 06-NOV-2001 (first entry)

XX  
 DE Human polynucleotide SEQ ID NO 11695.

XX  
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorders; arthritis; inflammation; ss.

XX  
 OS Homo sapiens.

XX  
 PN WO200164835-A2.

XX  
 PD 07-SEP-2001.

XX  
 PF 26-FEB-2001; 2001WO-US04927.

XX  
 PR 28-FEB-2000; 2000US-0515126.

XX  
 PR 18-MAY-2000; 2000US-0577409.



```

PF 22-OCT-1999; 99WO-US24046.
XX
PR 23-OCT-1998; 98US-0155209.
XX
PA (INCY-) INCYTE PHARM INC.
XX
PI Tang YT, Corley NC, Guegler KJ, Patterson C;
XX
DR WPI: 2000-350599/30.
DR P-PSDB; AAY94528.
XX
PT Purified polypeptide used for treating or preventing a disorder
PT characterized by expression or activity of lysine-rich statherin
PT proteins -
XX
PS Disclosure; Page 72; 75pp; English.
XX
CC The present invention relates to human lysine-rich statherin protein
CC (LRSP)(AAY94526). The cDNA sequence encoding this protein was identified
CC through analysis of a cDNA library of breast tumour tissue (BRISN0114).
CC The LRSP sequence was found to have homology with human statherin
CC protein (AAY9452) and human basic histidine-rich protein (the present
CC sequence). Human statherin is a phosphoprotein that acts as an inhibitor
CC of precipitation of calcium phosphate salts in the oral cavity. The LRSP
CC polypeptide and its antagonists may be useful for treating or preventing
CC disorders associated with the activity of LRSP. Such disorders include
CC autoimmune/inflammatory disorders (for example AIDS, allergies, asthma,
CC diabetes mellitus), bacterial and fungal infection and cancers (such as
CC leukemia, adenocarcinoma, melanoma). Antibodies to LRSP may be useful
CC for diagnosis of the above disorders.
XX
SQ Sequence 491 BP; 151 A; 90 C; 87 G; 163 T; 0 other;

Query Match 5.9%; Score 78.8; DB 21; Length 491;
Best Local Similarity 60.6%; Pred. No. 6.5e-10;
Matches 146; Conservative 0; Mismatches 93; Indels 2; Gaps 1;

QY 1093 CACAACCATACCAACCAATACCAATATACGTTTAAATATCATCAGTAACGCGAG 1152
DB 153 CACATCGAGGCTATAGATCAAAATATCTGTATGACAAATGATATCTTCAGTAAATCATGGG 212

QY 1153 ACATGATTATGAGGCTGATGGCAATACGACTTCTACATCCATATTCATCTTCA 1212
DB 213 GCATGATTATGAGGCTGATGGCAATATCGCTTTGGACTCGTGTATCTCATTTGCA 272

QY 1213 TACCATATACACTACTACCTTTTGTGNA--TCATCTAAGAGCAATGCGAATGAA 1270
DB 273 TACCGCATCACACTACTACTGCTTTTGAAGGAATATCATAAAGCAATGCGAATAAAA 332

QY 1271 AACCCATAAATTTACTGGATACCTTTGTTCCAGATCTTGCCTTTTCCATGTCATT 1330
DB 333 GAATACCATGATTTAGTGAATTCGTGTTTCAGGATACCTTCCCTTCCATATTCATT 392

QY 1331 G 1331
DB 393 G 393

RESULT 8
AAT24320
ID AAT24320 standard; cDNA to mRNA; 126 BP.
XX
AC AAT24320;
XX
DT 22-SEP-1996 (first entry)
DE Human gene signature HUMGS06347.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.

Query Match 5.9%; Score 78.6; DB 16; Length 126;
Best Local Similarity 86.6%; Pred. No. 4.6e-10;
Matches 84; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 859 CAACAGTAGCAAAATGTGTTTTCATAAAGTGGGAAGAAACAGCATTTTAAAGTAACTTTT 918
DB 14 CAAGAGCAGCAAAATGTGTTTTCATCACTGAGTGGGAAGAAAGCAGCAATTTAAATACNCTTT 73

QY 919 TGGGAGACTGATTTGAGTAATAAATAACTCTGGTCT 955
DB 74 TGGGAGACTGATTTGAGTAATAAATAACTCTGGTCT 110

RESULT 9
AAT23140
ID AAT23140 standard; cDNA to mRNA; 90 BP.
XX
AC AAT23140;
XX
DT 23-AUG-1996 (first entry)
DE Human gene signature HUMGS04875.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
XX
PN WO9514772-A1.
XX
PD 01-JUN-1995.
XX
PF 11-NOV-1994; 94WO-JP01916.
XX
PR 12-NOV-1993; 93JP-0355504.
XX
PA (MATS/) MATSUBARA K.
PA (OKUB/) OKUBO K.
XX
PI Matsubara K, Okubo K;
XX
DR WPI: 1995-206931/27.
XX
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
XX
PS Claim 1; Page 1582; 2245pp; Japanese.
XX
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in AAT19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
XX
SQ Sequence 126 BP; 44 A; 16 C; 25 G; 34 T; 7 other;

Query Match 5.9%; Score 78.6; DB 16; Length 126;
Best Local Similarity 86.6%; Pred. No. 4.6e-10;
Matches 84; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 859 CAACAGTAGCAAAATGTGTTTTCATAAAGTGGGAAGAAACAGCATTTTAAAGTAACTTTT 918
DB 14 CAAGAGCAGCAAAATGTGTTTTCATCACTGAGTGGGAAGAAAGCAGCAATTTAAATACNCTTT 73

QY 919 TGGGAGACTGATTTGAGTAATAAATAACTCTGGTCT 955
DB 74 TGGGAGACTGATTTGAGTAATAAATAACTCTGGTCT 110

RESULT 9
AAT23140
ID AAT23140 standard; cDNA to mRNA; 90 BP.
XX
AC AAT23140;
XX
DT 23-AUG-1996 (first entry)
DE Human gene signature HUMGS04875.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
XX
PN WO9514772-A1.
XX
PD 01-JUN-1995.

```

```

XX 11-NOV-1994; 94WO-JP01916.
XX 12-NOV-1993; 93JP-0355504.
XX (MATS/) MATSUBARA K.
XX (OKUB/) OKUBO K.
XX Matsubara K, Okubo K;
XX WPI; 1995-206931/27.
XX Identifying gene signatures in 3'-directed human cDNA library - e.g.
XX for diagnosis of abnormal cell function, by preparing cDNA that
XX reflects relative abundance of corresp. mRNA in specific human
XX tissues
XX Claim 1; Page 1297; 2245pp; Japanese.
XX A single-stranded DNA (or its complementary strand or the corresp.
XX double-stranded DNA) which comprises one of the 7837 "GS" sequences
XX given in AAT19001-T26837 and which is able to hybridise to part of
XX human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
XX sequences were obtained from 3'-directed cDNA libraries prepared
XX from various human tissues; synthesis of cDNA was initiated from the
XX 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
XX untranslated sequence is unique to a particular mRNA species, almost
XX all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
XX is constructed so as to reflect accurately the relative abundance of
XX different mRNAs in the particular tissue from which it was derived.
XX The appearance frequency of a given GS in a cDNA library can be
XX determined (esp. using primers and probes derived from the GS
XX sequences) as a means of diagnosing abnormal cell function or for
XX recognising different cell types.
XX Sequence 90 BP; 30 A; 12 C; 17 G; 31 T; 0 other;
XX
Query Match 5.7%; Score 76.4; DB 16; Length 90;
Best Local Similarity 97.8%; Pred. No. 1.4e-09;
Matches 88; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
XX
QY 392 GATCTTTTCATGTGCCTAATGCCATTTATTTGGTAAATGCC-GTTATTTGGTAAATACAGCAT 450
DB 1 GATCTTTTCATGTGCCTAATGCCATTTATTTGGTAAATGCCGTTATTTGGTAAATACAGCAT 60
QY 451 AGTTAAATAAAGTGTACAGTAAGTAATCTACA 480
DB 61 AGTTAAATAAAGTGTACAGTAAGTAATCTAAA 90
XX
RESULT 10
ABL54353
ID ABL54353 standard; DNA; 9881 BP.
XX
AC ABL54353;
XX
XX 29-JUL-2002 (first entry)
XX
DE Chemically treated apoptosis gene #27.
XX
KW Apoptosis; HIV; Bloom syndrome; cardiopathy;
KW neurodegenerative disorder; Herpes simplex virus; renal ischaemia;
KW amyotrophic lateral sclerosis; cancer; ds.
XX
OS Unidentified.
XX
XX WO200177164-A2.
XX
XX 18-OCT-2001.
XX
XX 06-APR-2001; 2001WO-EP03969.
XX
XX 06-APR-2000; 2000DE-1019058.
XX

```

```

PR 07-APR-2000; 2000DE-1019173.
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
XX (EPIG-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
XX
XX WPI; 2002-017444/02.
XX
XX Chemically modified sequences of genes associated with apoptosis are
XX useful to determine methylation patterns of genomic DNA samples for
XX diagnosis of associated diseases such as cancer.
XX Claim 1; Seq ID #53; 24pp; English.
XX
XX This invention relates to chemically pre-treated DNA of genes
XX associated with apoptosis. The nucleic acids are used to allocate
XX patients for specific therapy for HIV infection, Bloom syndrome,
XX cardiopathy, aging, neurodegenerative disorders, Herpes simplex virus
XX infection, renal ischaemia, amyotrophic lateral sclerosis, solid tumours
XX and cancers. This nucleotide sequence represents a chemically
XX treated apoptosis gene. Even SEQ ID numbers are the complementary
XX DNA strands to the odd SEQ ID numbers. The sequence data for this
XX patent is not represented in the printed specification but is based on
XX information supplied by the European patent office.
XX
XX Sequence 9881 BP; 2606 A; 196 C; 2308 G; 4771 T; 0 other;
XX
Query Match 5.6%; Score 74.6; DB 24; Length 9881;
Best Local Similarity 78.8%; Pred. No. 2e-08;
Matches 89; Conservative 0; Mismatches 24; Indels 0; Gaps 0;
XX
QY 860 AACAGCTAGCAATGTGGTTTCATAAAGTGGGAAGAAACAGCATTTTAAAGTAACCTTTT 919
DB 5771 AAGAGTAGTAATGTGGTTTATTAAGTGGGAAGAAAGTAGTAAATTTAAATAATTTT 5930
QY 920 GGGAGACTGATTTGGTAAATAATAAAGTCTGGTCTCCCTTAAGAAAAA 972
DB 5831 GGGAGATTGAATTGAGTAATAATAAATTTTATGTTTTCGTTAATAATAATA 5883
XX
RESULT 11
AAV89525
ID AAV89525 standard; cDNA; 516 BP.
XX
AC AAV89525;
XX
XX 15-FEB-1999 (first entry)
XX
DE EST clone CP289.
XX
XX Human; secreted protein; expressed sequence tag; EST; haematopoiesis;
XX tissue growth; activin; inhibitor; chemotaxis; chemokinesis; haemostatic;
XX receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;
XX gene therapy; ss.
XX
XX Homo sapiens.
XX
XX WO9845436-A2.
XX
XX 15-OCT-1998.
XX
XX 10-APR-1998; 98WO-US06955.
XX
XX 10-APR-1997; 97US-0838821.
XX
XX (GEMV ) GENETICS INST INC.
XX
XX Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
XX Racie LA, Spaulding V, Treacy M;
XX
XX WPI; 1999-070077/06.
XX

```

XX New polynucleotides encoding human secreted proteins - derived from  
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,  
PT ovary, pituitary, retina and colon cDNA libraries.  
XX  
PS Claim 1; Page 245; 618pp; English.  
XX  
CC The present sequence represents a human expressed sequence tag (EST).  
CC The polynucleotide, which is a secreted EST, and the encoded protein  
CC are predicted to have useful biological activities which would make  
CC them suitable for treating, preventing or ameliorating medical  
CC conditions in humans and animals, although no supporting data is  
CC given. Suggested activities include nutritional activity, immune  
CC stimulating or suppressing activity, haematopoiesis regulating  
CC activity, tissue growth activity, activin/inhibin activity,  
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic  
CC activity, receptor/ligand activity, anti-inflammatory activity,  
CC cadherin/tumour invasion suppressor activity, tumour inhibition  
CC activity. The polynucleotide may also be useful for gene therapy.  
XX  
SQ Sequence 516 BP; 156 A; 181 C; 97 G; 82 T; 0 other;  
  
Query Match 4.5%; Score 60.2; DB 20; Length 516;  
Best Local Similarity 89.0%; Pred. No. 2.9e-05;  
Matches 65; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
  
QY 1047 TATACCAAGAGCGCTATCAGCCAGTTCAGACACCACTATACGACACCAATACCAA 1106  
DB 139 TATGGGTATGGCCCTTATCAGCCAGTTCAGACACCACTATACGACACCAATACCAA 198  
QY 1107 CCACATATACCAAC 1119  
DB 199 CCACATATACCAAC 211  
  
RESULT 12  
ABN43273  
ID ABN43273 standard; DNA; 60 BP.  
AC ABN43273;  
XX  
XX 15-JUL-2002 (first entry)  
XX  
DE Human spliced transcript detection oligonucleotide SEQ ID NO:16021.  
XX  
KW Human; mouse; rat; splice transcript; detection; RNA transcript;  
KW splice variant; transcriptome; oligonucleotide library; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200210449-A2.  
XX  
PD 07-FEB-2002.  
XX  
XX 20-JUL-2001; 2001WO-IB01903.  
XX  
XX 28-JUL-2000; 2000US-221607P.  
PR 02-MAY-2001; 2001US-287724P.  
XX  
XX (COMP-) COMPUGEN INC.  
PA  
PI Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;  
XX  
XX WPI; 2002-257383/30.  
XX  
XX New oligonucleotide libraries comprising oligonucleotides which  
XX selectively hybridize to mRNAs transcribed from a transcription unit of  
PT a genome, useful for detecting tissue-, pathology-, and  
PT developmental-specific genes.  
XX  
XX Example 1; SEQ ID 16021; 47pp; English.  
PS  
XX The present invention describes oligonucleotide libraries for detecting

CC messenger RNAs that populate a (sub-)transcriptome, where the  
CC (sub-)transcriptome comprises messenger RNAs transcribed from multiple  
CC transcription units that populate a genome. The library comprises  
CC several oligonucleotides, each capable of hybridising selectively to a  
CC set of messenger RNAs transcribed from a given transcription unit of  
CC the genome, which encodes one or more messenger RNA splice variants.  
CC The oligonucleotide libraries are useful for detecting mRNAs from a  
CC biological sample, in expression profiling studies, in qualitatively or  
CC quantitatively characterising the corresponding transcriptome, and in  
CC detecting RNA transcripts and splice variants of human or animal  
CC transcriptomes. The libraries may also be used as specialised mini  
CC libraries to detect transcripts of a sub-transcriptome under a  
CC particular biological or pathological state, and so allowing the  
CC detection of tissue- and pathology-specific genes such as those genes  
CC only expressed in specific tissue under a specific pathological  
CC condition; to detect developmental specific genes; and to detect RNA  
CC transcripts and splice variants of a transcriptome of a patient suffering  
CC from a particular disorder. ABN27253 to ABN59589 represent  
CC oligonucleotide sequences from rats, humans and mice, which are used in  
CC the exemplification of the present invention.  
CC N.B. The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 60 BP; 22 A; 25 C; 4 G; 9 T; 0 other;

Query Match 4.2%; Score 56.4; DB 24; Length 60;  
Best Local Similarity 98.3%; Pred. No. 0.00012;  
Matches 57; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1059 CCTATCAGCCAGTTCAGACACCACTATACGACACCAATACCAATACC 1116  
DB 3 CCTATCAGCCAGTTCAGACACCACTATACGACACCAATACCAATACC 60

RESULT 13  
ABL54354/C  
ID ABL54354 standard; DNA; 9881 BP.  
XX  
XX ABL54354;  
XX  
XX 29-JUL-2002 (first entry)  
XX  
DE Chemically treated apoptosis gene complementary to gene #27.  
XX  
XX Apoptosis; HIV; Bloom syndrome; cardiopathy;  
KW neurodegenerative disorder; Herpes simplex virus; renal ischaemia;  
KW amyotrophic lateral sclerosis; cancer; ds.  
XX  
XX Unidentified.  
XX  
XX WO200177164-A2.  
XX  
XX 18-OCT-2001.  
XX  
XX 06-APR-2001; 2001WO-EP03969.  
PF  
XX 06-APR-2000; 2000DE-1019058.  
PR 07-APR-2000; 2000DE-1019173.  
PR 30-JUN-2000; 2000DE-1032529.  
PR 01-SEP-2000; 2000DE-1043826.  
XX  
XX (EPIG-) EPIGENOMICS AG.  
XX  
XX Olek A, Piepenbrock C, Berlin K;  
PI  
XX WPI; 2002-017444/02.  
DR  
XX Chemically modified sequences of genes associated with apoptosis are  
PT useful to determine methylation patterns of genomic DNA samples for  
PT diagnosis of associated diseases such as cancer.  
XX  
XX Claim 1; Seq ID #54; 24pp; English.

XX This invention relates to chemically pre-treated DNA of genes  
 CC associated with apoptosis. The nucleic acids are used to allocate  
 CC patients for specific therapy for HIV infection, Bloom syndrome,  
 CC cardiopathy, aging, neurodegenerative disorders, Herpes simplex virus  
 CC infection, renal ischaemia, amyotrophic lateral sclerosis, solid tumours  
 CC and cancers. This nucleotide sequence represents a chemically  
 CC treated apoptosis gene. Even SEQ ID numbers are the complementary  
 CC DNA strands to the odd SEQ ID numbers. The sequence data for this  
 CC patent is not represented in the printed specification but is based on  
 CC information supplied by the European patent office.

XX  
 SQ Sequence 9881 BP; 2874 A; 196 C; 2093 G; 4718 T; 0 other;  
 Query Match 4.2%; Score 55.8; DB 24; Length 9881;  
 Best Local Similarity 67.1%; Pred. No. 0.001;  
 Matches 96; Conservative 0; Mismatches 42; Indels 5; Gaps 1;  
 QY 830 CAATATTTCTGATCAGATAGTCCCTGTCACAGTAGCAAAATGTTTCATATAAGTGG 889  
 DB 4136 CAATATTTCTGATCAGATAGTCCCTGTCACAGTAGCAAAATGTTTCATATAAGTGG 889  
 QY 890 GAAGAAACAGCAGATTATAAGTAACTTTTGGGAGACTGATTTGAGTAATAATAAACTT 4022  
 DB 4081 AAAAAAACAACAATTTAAATFARCTTTTAAAAAACTAAATTAATAATAATAAACTT 4022  
 QY 950 TGCTCTCCCTTAGAAAAA 972  
 DB 4021 CAATCTTTGCTAATAATAATA 3999

RESULT 14  
 ABL33026  
 ID ABL33026 standard; DNA; 6056 BP.  
 AC ABL33026;  
 XX  
 DT 26-MAR-2002 (first entry)  
 DE Human immune system associated gene SEQ ID NO: 999.  
 KW Human; immune system disease; cytosine methylation; antiasthmatic;  
 KW antiarteriosclerotic; antianaemic; cytosine; neotropic;  
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
 KW gene; ds.

OS Homo sapiens.  
 XX  
 PN WO2000200928-A2.  
 XX  
 PD 03-JAN-2002.  
 XX  
 PF 02-JUL-2001; 2001WO-EP07537.  
 XX  
 PR 30-JUN-2000; 2000DE-1032529.  
 PR 01-SEP-2000; 2000DE-1043826.  
 XX  
 PA (EPIG-) EPIGENOMICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K;  
 XX  
 DR WPI; 2002-130909/17.  
 XX  
 PT Nucleic acid comprising fragment of chemically modified gene, useful  
 PT for diagnosis and treatment of diseases associated with abnormal  
 PT cytosine methylation  
 XX  
 PS Claim 1; SEQ ID NO 999; 32pp + Sequence Listing; German.  
 XX

CC The present invention provides a number of human immune system associated  
 CC genes which are modified by the methylation of cytosines. The sequences  
 CC can be used in the diagnosis and treatment of immune system disorders,  
 CC including eye diseases such as retinopathy, neovascular glaucoma and  
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
 CC diseases. The present sequence is a gene of the invention.

XX  
 SQ Sequence 6056 BP; 1780 A; 54 C; 1079 G; 3143 T; 0 other;

Query Match 3.6%; Score 48.4; DB 24; Length 6056;  
 Best Local Similarity 45.9%; Pred. No. 0.06;  
 Matches 166; Conservative 0; Mismatches 196; Indels 0; Gaps 0;  
 QY 688 TTCTTTGTTGTTTAAAGACAGAGACTCTGTCCTCAAAAAAGGACATTTATCATT 747  
 DB 3741 TTTTATGATTTTCGAAATGAAAAATTTTATTTGTTTTTAAATAATAATAATA 3800  
 QY 748 ATAACATCTTATAGAGCCCCCTAAATTTCTATCTGAAGGACACGTTTTTTTTTAAACA 807  
 DB 3801 AGTAGGTTTATGTAAGTATTTTGTGTTTTGATTTTAGGAATTTTGTGTTTTATATTA 3860  
 QY 808 GTTAAGTACTGATGTCAACAGACAATATTTCTGATCAGATAGTCCCTCTCAACAGTAG 867  
 DB 3861 ATTTGCTATTTGTAATAATAATAATAATAATAATAATAATAATAATAATAATA 3920  
 QY 868 CAAATGTTGTTTCAATAAGTGGGAAGAAACAGACATTTTAAAGTAACTTTTGGGAGACT 927  
 DB 3921 GTAATATAGTTAGTATATGTTTAAAGAAAAAATGTTTAAAGAAATTTCTTTTATTAAT 3980  
 QY 928 GATTGAGCTAATAATAAACTCTGGTCTCCCTTAAGAAAAAACCCTTCACCTTTAC 987  
 DB 3981 TATTTTTAAGATGTGATATTTATTTGTTTTTAAACGTTATATTTTATAAGATTAT 4040  
 QY 988 TGTGCTATTATATCCCTTAGTCCCAAGTTAATTTCTGATATTTCTGATATTTGCTTTT 1047  
 DB 4041 TTTTAAATGTTGAATATAGTTTAAATATATATGATTTTAAATTTGTTTAAATAT 4100  
 QY 1048 AT 1049  
 DB 4101 TT 4102

RESULT 15  
 AAS46787/C  
 ID AAS46787 standard; DNA; 61020 BP.  
 XX  
 AC AAS46787;  
 XX  
 DT 18-DEC-2001 (first entry)  
 DE Tumour suppressor gene derived chemically modified sequence #513.  
 DE Human; tumour suppressor gene; oncogene; antitumour; cytostatic;  
 KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;  
 KW cytosine methylation; ds.

OS Homo sapiens.  
 XX  
 PN WO200168912-A2.  
 XX  
 PD 20-SEP-2001.  
 XX  
 PF 15-MAR-2001; 2001WO-EP02955.  
 XX  
 PR 15-MAR-2000; 2000DE-1013847.  
 PR 06-APR-2000; 2000DE-1019058.  
 PR 07-APR-2000; 2000DE-1019173.  
 PR 30-JUN-2000; 2000DE-1032529.  
 PR 01-SEP-2000; 2000DE-1043826.  
 XX  
 PA (EPIG-) EPIGENOMICS AG.





GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 14, 2003, 21:15:00 ; Search time 97 Seconds  
(without alignments)  
4208.111 Million cell updates/sec

Title: US-09-830-244B-2

Perfect score: 1331

Sequence: 1 ctagtgttttagaatcaag.....gcctttccaatgcaacttg 1331

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA.\*

- 1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*
- 2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*
- 3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*
- 4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*
- 5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq.\*
- 6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39.6	3.0	7218	1	US-08-232-463-14
2	38.6	2.9	6265	4	US-09-129-112-3
3	38	2.9	588	4	US-09-385-982-133
4	37.2	2.8	5852	1	US-07-867-106-2
5	37.2	2.8	246240	2	US-08-724-394A-20
6	37.2	2.8	246240	2	US-08-724-394A-21
7	37.2	2.8	246240	2	US-08-724-394A-22
8	36.4	2.7	12730	4	US-09-004-838-91
9	36.4	2.7	12793	4	US-09-004-838-124
10	36.2	2.7	762	4	US-09-134-001C-748
11	36.2	2.7	8302	4	US-09-234-827B-1
12	36	2.7	3718	4	US-09-424-283-6
13	35.8	2.7	72928	3	US-09-009-913-1
14	35.6	2.7	3780	4	US-09-134-001C-920
15	35.6	2.7	19124	2	US-08-487-826B-13
16	35	2.6	1474	4	US-08-821-994-64
17	35	2.6	2192	3	US-08-714-918-14
18	35	2.6	2192	4	US-09-265-315-14
19	35	2.6	2192	4	US-09-265-315-14
20	35	2.6	2192	4	US-09-266-417-14
21	35	2.6	9048	3	US-08-973-273-4
22	35	2.6	36551	4	US-09-738-894A-3
23	34.8	2.6	4517	4	US-09-140-804-9
24	34.8	2.6	11056	4	US-09-004-838-23
25	34.8	2.6	15062	4	US-09-004-838-89
26	34.8	2.6	62804	4	US-09-800-960-3
27	34.4	2.6	2173	6	5168051-10
					Patent No. 5168051

28	34.2	2.6	688	4	US-08-998-416-915	Sequence 915, App
29	34.2	2.6	1394	4	US-09-247-155-76	Sequence 76, Appl
C 30	34.2	2.6	1408	3	US-08-889-841B-22	Sequence 22, Appl
C 31	34.2	2.6	7228	2	US-08-850-049-128	Sequence 128, App
C 32	34.2	2.6	7228	2	US-08-850-049-129	Sequence 129, App
C 33	34.2	2.6	7228	2	US-08-050-478-128	Sequence 128, App
C 34	34.2	2.6	7228	2	US-08-050-478-129	Sequence 129, App
C 35	34.2	2.6	7228	2	US-08-050-478-129	Sequence 128, App
C 36	34.2	2.6	7228	4	US-09-414-117-128	Sequence 128, App
C 37	34.2	2.6	7228	4	US-09-414-117-128	Sequence 129, App
C 38	34.2	2.6	7228	4	US-09-678-437-128	Sequence 129, App
C 39	34.2	2.6	7228	4	US-09-678-437-129	Sequence 128, App
C 40	34.2	2.6	15581	3	US-08-646-538-35	Sequence 35, Appl
C 41	34	2.6	1744	4	US-09-503-222-35	Sequence 35, Appl
C 42	34	2.6	1745	4	US-09-511-625B-19	Sequence 19, Appl
C 43	34	2.6	32768	4	US-08-961-527-71	Sequence 71, Appl
C 44	33.8	2.5	1786	1	US-07-920-430-19	Sequence 19, Appl
C 45	33.8	2.5	1786	1	US-08-066-299-9	Sequence 9, Appl

## ALIGNMENTS

RESULT 1  
US-08-232-463-14  
; Sequence 14 Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION NUMBER: US/08/232,463  
; APPLICATION NUMBER: US/08/232,463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/995,313  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)836-9300  
; TELEFAX: (703)683-4109  
; TELEX: 899149  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7218 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; CLONE: pTZgpt-Fls  
US-08-232-463-14

Query Match 3.0%; Score 39.6; DB 1; Length 7218;

		Best Local Similarity	7.0%; Pred. No. 0.54;	
Matches	18;	Conservative	138; Mismatches	102; Indels 0; Gaps 0.
QY	545	GCATGCTCGAGAAAGACGTACTAATATTTTTGGTAGCTTTATCTGAATTCAGAATGCC	604	
Db	1037	GCTTGCGTCAGGTCGAGGAGCCTTCGCAATTTTTTTTTTTTTTTTTTTTTTTTTTTT	1096	
QY	605	TGCTTCCCCTGCAGGTGTGTTTCCTCTTAGCATCCCTCATTTGAATCCCCCTGGAGCAC	664	
Db	1097	YYY	1156	
QY	665	AGCACAGTTAGTAGAACCTCCATTTCTGTTTGGTTTTTAAGACAGACTCTGTCT	724	
Db	1157	YYY	1216	
QY	725	CRAAAAAAGGACATTTATATAACATCTTATTAGAGCCCTAATTTCTTATCTGAA	784	
Db	1217	YYY	1276	
QY	785	GGCACTGTTTTTTTTTTT	802	
Db	1277	XXXXXXXXXXXXXXX	1294	

```

RESULT 2
US-09-129-112-3/c
; Sequence 3, Application US/09129112
; Patent No. 6465716
; GENERAL INFORMATION:
; APPLICANT: Etzler, Marilynn E.
; APPLICANT: Murphy, Judith B.
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: A No. 6465716 Factor Binding Protein From Legume Roots
; FILE REFERENCE: 023070-0798100S
; CURRENT APPLICATION NUMBER: US/09/129,112
; CURRENT FILING DATE: 1998-08-04
; PRIOR APPLICATION NUMBER: US 08/907,226
; PRIOR FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 6265
; TYPE: DNA
; ORGANISM: Dolichos biflorus
; FEATURE:
; OTHER INFORMATION: genomic sequence of NBP46 (DB46)
; NAME/KEY: exon
; LOCATION: (633)..(944)
; NAME/KEY: Intron
; LOCATION: (945)..(1022)
; NAME/KEY: exon
; LOCATION: (1023)..(1151)
; NAME/KEY: Intron
; LOCATION: (1152)..(1559)
; NAME/KEY: exon
; LOCATION: (1560)..(1616)
; NAME/KEY: Intron
; LOCATION: (1617)..(1697)
; NAME/KEY: exon
; LOCATION: (1698)..(1790)
US-09-129-112-3

```

QY	846	GATAGTCCCTGTCACACAGTAGCAAAATGTGGTTCATAAAGTGGGAAGAAAACAGCATTT	905
Db	2670	AAAATAATAATTTATAAATGTATCAGTTATGCTATTTTATATTATCAATTAAGCAGACAAT	2611
QY	906	TAAAGTAACCTTTTGGGAGACTGATTTGAGTAATAATAAAACTCTGGTCCCTTAAAGAA	965
Db	2610	TTTATTAAATAATATAATTTGAATTTAAGAAATCCCAACTACTAGTTTGTGAAAACATAA	2551
QY	966	AAAAAACCCCTCCACCTTTACTGTGTCATATATATCCCTTAGTTCCAAAGTTAATTAT	1025
Db	2550	CAAAGAACCTAATTTCTCTTTTGGAAACATATCCCTCATGSGTTTCATCCACTTTAATTTC	2491
QY	1026	CTTATTTCGGGATATGCTTTTATACCAAGAGCGCTTATCAGCCAGTTCAGAACCAACCA	1085
Db	2490	ATCATTCGCTCTCTCTTTCTTTTGTGGCAATCTTTTAGTGAAGATTAGTCAATTG	2431
QY	1086	CTATAGCCACAAACCATACCAACCAATACCAACAATATACGTTTAAATATCATCAGTAA	1145
Db	2430	CGCGACCCTAAAAGAAACACACAAATAAAATGAATAGTGAATTTAATTACTTTATACAA	2371
QY	1146	CTCAGGACATGATTATTGAGGCTTGATGGCGAAATACGACTTCTACATCCATTTCTCA	1205
Db	2370	AATTATATCGTCGTGATTCAATTCGACATAAATAAAAAATCATATTTAGAGATAGTGTG	2311
QY	1206	TCCTTT 1210	
Db	2310	CATTT 2306	

```

RESULT 3
US-09-385-982-133
; Sequence 133, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; TITLE OF INVENTION: PRODUCTS: I
; FILE REFERENCE: CCDNA-260X
; CURRENT APPLICATION NUMBER: US/09/385,982
; CURRENT FILING DATE: 1998-08-30
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1998-06-08
; EARLIER APPLICATION NUMBER: 60/117,393
; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,639
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 133
; LENGTH: 588
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(588)
; OTHER INFORMATION: n = A,T,C or G
US-09-385-982-133

```

	Query Match	2.9%	Score 38;	DB 4;	Length 588;
	Best Local Similarity	41.3%;	Pred. No. 0.57;		
	Matches	95;	Conservative	Mismatches 135;	Indels 0; Gaps 0;
QY	35	TGTCATGTACCAACGTGAATTTACAGTGTTTACAAATGCTCTGGAAATTTTGCACATGC	94		
Db	96	TGGCGNAGGTAACACAGGTCAGTATTGGTCNNGTGACAAGAGNACGCAANTCTGCCNGA	155		
QY	95	CATAGGGAATGTTAAGTTTACTTTCGGCTGGAAATTTATCAGACTTTGTCAGTAAACAAGTTGA	154		
Ddb	156	CANTGANGTNAANAAGGTTTATNTTTTACANTTATNTNNANATATNTNNNNAANTATTAA	215		
QY	155	AGTTTAGCAGATGAGGGGGGAATATTGAGGCCCTTAGGCTAACAAATAATTCAGTATCT	214		





; TELEPHONE: (415) 576-0200  
; TELEFAX: (415) 576-0300  
; INFORMATION FOR SEQ ID NO: 91:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 12730 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA  
; FEATURE:  
; NAME/KEY: -  
; LOCATION: 1..12730  
; OTHER INFORMATION: /note= "RG2C"  
US-09-004-838-91

Query Match 2.7%; Score 36.4; DB 4; Length 12730;  
Best Local Similarity 51.2%; Pred. No. 4.7;  
Matches 85; Conservative 0; Mismatches 81; Indels 0; Gaps 0;  
QY 678 GAACTCTCCATTCCTTTGTTGTTTAAAGACAGAGACTCTGTCTCAAAAAAGGAC 737  
Db ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||  
QY 3997 GAAATGACATTTAAAGTTGTTAACTTATAATCCGTTGCTTATTTCACATAAAGTAAT 3938  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 738 ATTATCATTAACATCTATTAGAGCCCTAAATTTCTATCTGAAGGCACGTGTTTT 797  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 3937 ATTACATTTAAGCATTTATTAGCTAAAGTTAGTAATTTCTAATTTAAATGCTTT 3878  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 798 TTTTAAACAGTTAAGTACTGTCTCAACAGACAAATATTTCTGAT 843  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 3877 AAATCTACTATTATAATGATCTCTAATGAAATCATGTTGAT 3832  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 9  
US-09-004-838-124/c  
; Sequence 124, Application US/09004838  
; Patent No. 6350933  
; GENERAL INFORMATION:  
; APPLICANT: Michelmore, Richard W.  
; APPLICANT: Shen, Kathy  
; APPLICANT: Meyers, Blake  
; TITLE OF INVENTION: Procedures and Materials for  
; TITLE OF INVENTION: Confering pest Resistance in Plants  
; NUMBER OF SEQUENCES: 140  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/004,838  
; FILING DATE: 09-JAN-1998  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/781,734  
; FILING DATE: 10-JAN-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Einhorn, Gregory P.  
; REGISTRATION NUMBER: 38,440  
; REFERENCE/DOCKET NUMBER: 023070-078810US  
; TELEPHONE: (415) 576-0200  
; TELEFAX: (415) 576-0300  
; INFORMATION FOR SEQ ID NO: 124:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 12793 base pairs  
; TYPE: nucleic acid

; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA  
; FEATURE:  
; NAME/KEY: -  
; LOCATION: 1..12793  
; OTHER INFORMATION: /note= "RG2S"  
US-09-004-838-124

Query Match 2.7%; Score 36.4; DB 4; Length 12793;  
Best Local Similarity 51.2%; Pred. No. 4.7;  
Matches 85; Conservative 0; Mismatches 81; Indels 0; Gaps 0;  
QY 678 GAACTCTCCATTCCTTTGTTGTTTAAAGACAGAGACTCTGTCTCAAAAAAGGAC 737  
Db ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||  
QY 738 ATTATCATTAACATCTATTAGAGCCCTAAATTTCTATCTGAAGGCACGTGTTTT 797  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 3789 ATTACATTTAAGCATTTATTAGCTAAAGTTAGTAATTTCTAATTTAAATGCTTT 3730  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 798 TTTTAAACAGTTAAGTACTGTCTCAACAGACAAATATTTCTGAT 843  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 3729 AAATCTACTATTATAATGATCTCTAATGAAATCATGTTGAT 3684  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 10  
US-09-134-001C-748/c  
; Sequence 748, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 748  
; LENGTH: 762  
; TYPE: DNA  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-748

Query Match 2.7%; Score 36.2; DB 4; Length 762;  
Best Local Similarity 54.0%; Pred. No. 1.9;  
Matches 74; Conservative 0; Mismatches 53; Indels 0; Gaps 0;  
QY 586 CATTCTCTTTGTTGTTTAAAGACAGAGACTCTGTCTCAAAAAAGGACATTTATCA 745  
Db ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||  
QY 181 CGTTTATTGAAATAATTTTACATTTGAATACATTTTTCATCTGAAAGTGACATAATAT 122  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 746 TTATAACATCTTATTAGAGCCCTAAATTTCTATCTGAAGGCACGTGTTTTTTTAA 805  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 121 TTGTCTCTTCTTATAGCGCATACATTCACGATATGATAGTCTCTTTGTTTCAA 62  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 806 CAGTTAAGTACTGATGT 822  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 61 TAAAAAGACACTGATTT 45  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 11  
US-09-234-827B-1/c  
; Sequence 1, Application US/09234827B  
; Patent No. 6448471  
; GENERAL INFORMATION:  
; APPLICANT: Fuzio, Piotr S.  
; APPLICANT: Grundler, Florian M.W.  
; TITLE OF INVENTION: Nematode feeding structure specific gene and its

; TITLE OF INVENTION: application to produce nematode resistant plants  
 ; FILE REFERENCE: U-012084-2  
 ; CURRENT APPLICATION NUMBER: US/09/234,827B  
 ; CURRENT FILING DATE: 1999-01-21  
 ; PRIOR APPLICATION NUMBER: US 60/072,142  
 ; PRIOR FILING DATE: 1998-01-22  
 ; NUMBER OF SEQ ID NOS: 32  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 1  
 ; LENGTH: 8302  
 ; TYPE: DNA  
 ; ORGANISM: Arabidopsis thaliana  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (4038)..(4256)  
 ; NAME/KEY: CDS  
 ; LOCATION: (4807)..(5604)  
 ; NAME/KEY: CDS  
 ; LOCATION: (6777)..(6827)  
 ; NAME/KEY: CDS  
 ; LOCATION: (6919)..(7935)  
 ; US-09-234-827B-1

Query Match 2.7%; Score 36.2; DB 4; Length 8302;  
 Best Local Similarity 51.2%; Pred. No. 4.5;  
 Matches 111; Conservative 0; Mismatches 103; Indels 3; Gaps 1;

QY 683 CTCACATTCCTTTGTTTTTAAACAGACAGACTGCTCTCAAAAAGGACATTTA 742  
 DB 805 CTTCTGTTGTTGTTTTTATAGACAGTCCGAAACACATCAATCACTTA 746  
 QY 743 TCATTAAACATCTATTAGAGCCCTTAATTTCTATCTGAAGCAGCTGTTTTTTT 802  
 DB 745 ATTTTAAATCAATATATACATTTAAATTCGTGTA---TTATTTCTAAATTA 689  
 QY 803 AAACAGTTAGTACTGATGACACAGACAAATATTTCTGATCAGATGCCCCGTGCAAC 862  
 DB 688 AATAATTAATATTAATGTCCTCAATATAAATGATCGTGTGTTTTATTGGGTTTTACC 629  
 QY 863 AGTAGCAATGTTGTTTCATTAAGTGGGAAGAAACA 899  
 DB 628 TGTAAGCACTCTGTTTTATCAAAAAGAAAAAAA 592

RESULT 12  
 US-09-424-283-6  
 ; Sequence 6, Application US/09424283  
 ; Patent No. 6437219  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Grimes, et al.  
 ; TITLE OF INVENTION: Sucrose binding proteins  
 ; FILE REFERENCE: 4630-50206  
 ; CURRENT APPLICATION NUMBER: US/09/424,283  
 ; CURRENT FILING DATE: 1999-11-19  
 ; PRIOR APPLICATION NUMBER: PCT/US98/10465  
 ; PRIOR FILING DATE: 1998-05-21  
 ; PRIOR APPLICATION NUMBER: US 60/047,568  
 ; PRIOR FILING DATE: 1997-05-22  
 ; NUMBER OF SEQ ID NOS: 15  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 6  
 ; LENGTH: 3718  
 ; TYPE: DNA  
 ; ORGANISM: Glycine max  
 ; US-09-424-283-6

Query Match 2.7%; Score 36; DB 4; Length 3718;  
 Best Local Similarity 47.0%; Pred. No. 3.8;  
 Matches 111; Conservative 0; Mismatches 125; Indels 0; Gaps 0;  
 QY 983 TTATCTGCTGCTATATATCCCTTAGTCCAAAGCTTAATATCTTATTTCTGGATATG 1042  
 DB 3257 TGTACGCATTAATTTGATTTTTTTCTCCACATTTAATGAGGTGAATCAGTTAGAA 3316

QY 1043 CTTTATACCAAGAGCCTTATCAGCCAGTTCACGAACACACACTATACGCACACACATA 1102  
 DB 3317 ATATTAAAAAATAAATAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3376  
 QY 1103 CCAACACCAATACCAACATATATACGTTTATATATCATCATCATCATCATCATCAT 1162  
 DB 3377 ACAATGAATAGAGACCAATTAGAACCATTTATTTCTTACAAATTTAAAGAAAAAGCTTTT 3436  
 QY 1163 TGAGGCTTCATGGCAAAATACGACTTCTACATCCATATATCTCATCTTCATACCAT 1218  
 DB 3437 TTAACAATATATACATTTATATATATATATATATATATATATATATATATATATAT 3492

RESULT 13  
 US-09-009-913-1  
 ; Sequence 1, Application US/09009913  
 ; Patent No. 6087485  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Axys Pharmaceuticals, Inc.  
 ; TITLE OF INVENTION: Asthma Related Genes  
 ; NUMBER OF SEQUENCES: 339  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Bozicevic & Reed, LLP  
 ; STREET: 285 Hamilton Ave, Suite 200  
 ; CITY: Palo Alto  
 ; STATE: CA  
 ; COUNTRY: USA  
 ; ZIP: 94301  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: FastSeq for Windows Version 2.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/009,913  
 ; FILING DATE: 21-JAN-1998  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER:  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Sherwood, Pamela J  
 ; REGISTRATION NUMBER: 36,677  
 ; REFERENCE/DOCKET NUMBER: SEQ-4P  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 650-327-3231  
 ; TELEFAX: 650-327-3231  
 ; TELEX:  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 72928 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: Genomic DNA  
 ; US-09-009-913-1

Query Match 2.7%; Score 35.8; DB 3; Length 72928;  
 Best Local Similarity 54.1%; Pred. No. 13;  
 Matches 73; Conservative 0; Mismatches 62; Indels 0; Gaps 0;  
 QY 708 AGACAGAGACTCTGCTCAAAAAGGACATTTATCATTAACATCTTATTAGAGCCC 767  
 DB 68237 AGAGTGAGACTCTGCTCAAAAATTAATAAAAAAACTCTGTAATTTACTTTTG 68296  
 QY 768 CTAATTTCTTCTGAAGGACGTGTTTTTTTTTAAACAGTTAACTACTGTGTCACAA 827  
 DB 68297 CACCACATAATATGATATACACATTTATTTTAAAGATATTTTGACATTTGCTTTTAA 68356  
 QY 828 GACAAATATTTCTGA 842  
 DB 68357 TATAAATTTTTTAA 68371

```

RESULT 14
US-09-134-001c-920/c
; Sequence 920, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134.001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 920
; LENGTH: 3780
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
; US-09-134-001c-920

```

[illegible]

RESULT 15  
US-08-487-826B-13/c  
; Sequence 13, Application US/08487826B  
; Patent No. 5993827  
; GENERAL INFORMATION:  
; APPLICANT: Sim, Kim L.  
; APPLICANT: Chitnis, Chetan  
; APPLICANT: Miller, Louis H.  
; APPLICANT: Peterson, David S.  
; APPLICANT: Su, Xin-zhaun  
; APPLICANT: Wellemis, Thomas E.  
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE  
; NUMBER OF SEQUENCES: 45  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Knobbe Martens Olson & Bear  
; STREET: 620 Newport Center Drive 16th Floor  
; CITY: Newport Beach  
; STATE: California  
; COUNTRY: US  
; ZIP: 92660  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/487.826B  
; FILING DATE: 10-SEP-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Israelsen, Ned



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 14, 2003, 21:17:40 ; Search time 134 Seconds  
(without alignments)  
8712.764 Million cell updates/sec

Title: US-09-830-244b-2

Perfect score: 1331

Sequence: 1 ctatgttttagaatacaag.....gccttttccaatgctacattg 1331

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 593429 seqs, 438583890 residues

Total number of hits satisfying chosen parameters: 1186858

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications.NA.\*

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*  
12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq.\*  
13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*  
14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	104.6	7.9	438	US-09-924-340-5	Sequence 5, Appli
2	104.6	7.9	438	US-09-924-340-5	Sequence 5, Appli
3	74.8	5.6	433	US-09-983-965-5821	Sequence 5821, Ap
4	46.8	3.5	3104	US-10-037-598-32	Sequence 32, Appl
5	46.8	3.5	513509	US-09-754-853A-4	Sequence 4, Appli
6	43.2	3.2	65359	US-09-804-472-3	Sequence 3, Appli
7	43.2	3.2	6055	US-10-091-504-1746	Sequence 1746, Ap
8	42.2	3.2	6055	US-09-764-869-1746	Sequence 1746, Ap
9	42.2	3.2	7809	US-10-091-504-1744	Sequence 1744, Ap
10	42.2	3.2	7809	US-09-764-869-1744	Sequence 1744, Ap
11	42.2	3.2	7809	US-09-764-869-1745	Sequence 1745, Ap
12	42.2	3.2	7809	US-09-764-869-1745	Sequence 1745, Ap
13	40.6	3.1	556	US-10-092-154-1685	Sequence 1685, Ap
14	40.6	3.1	556	US-09-764-847-1685	Sequence 1685, Ap
15	40.4	3.0	17849	US-10-092-154-1315	Sequence 1315, Ap
16	40.4	3.0	17849	US-09-764-847-1315	Sequence 1315, Ap
17	40.4	3.0	17862	US-10-092-154-1313	Sequence 1313, Ap
18	40.4	3.0	17862	US-09-764-847-1313	Sequence 1313, Ap
19	39.6	3.0	151	US-09-864-761-22646	Sequence 22646, A

20	39.6	3.0	492	10	US-09-864-761-5885	Sequence 5885, Ap
21	39.6	3.0	2000	9	US-09-938-842A-4305	Sequence 4305, Ap
22	39.6	3.0	44848	9	US-09-988-113-42	Sequence 42, Appli
23	39.6	3.0	44848	10	US-09-776-874A-42	Sequence 42, Appli
24	39	2.9	378361	9	US-09-901-136-3	Sequence 3, Appli
25	38.6	2.9	5236	10	US-09-925-297-363	Sequence 363, App
26	38.6	2.9	6265	10	US-09-129-112-3	Sequence 3, Appli
27	38.6	2.9	10758	12	US-10-044-090-61	Sequence 61, Appli
28	38.2	2.9	2000	9	US-09-938-842A-3631	Sequence 3631, Ap
29	38.2	2.9	2286	9	US-10-228-796-3	Sequence 3, Appli
30	38.2	2.9	2286	10	US-09-191-687B-3	Sequence 3, Appli
31	38	2.9	1282	9	US-09-938-842A-4177	Sequence 4177, Ap
32	38	2.9	33023	10	US-09-880-107-3350	Sequence 3350, Ap
33	37.4	2.8	13715	7	US-08-781-986A-195	Sequence 195, App
34	37.4	2.8	397658	10	US-09-813-320-3	Sequence 3, Appli
35	37.2	2.8	4000	9	US-09-981-353-8	Sequence 8, Appli
36	37	2.8	2000	9	US-09-938-842A-2721	Sequence 2721, Ap
37	36.8	2.8	32204	9	US-09-764-872-517	Sequence 517, App
38	36.8	2.8	32204	10	US-10-072-349-327	Sequence 327, App
39	36.8	2.8	32204	10	US-09-764-855-327	Sequence 327, App
40	36.6	2.7	495	9	US-10-092-154-219	Sequence 219, App
41	36.6	2.7	495	10	US-09-764-847-219	Sequence 219, App
42	36.6	2.7	1009	9	US-10-076-816-18	Sequence 18, Appli
43	36.6	2.7	2000	9	US-09-938-842A-2839	Sequence 2839, Ap
44	36.4	2.7	99014	10	US-09-880-107-3428	Sequence 3428, Ap
45	36.2	2.7	376	10	US-09-960-352-9302	Sequence 9302, Ap

#### ALIGNMENTS

#### RESULT 1

US-09-924-340-5  
; Sequence 5, Application US/09924340  
; Publication No. US20030027248A1  
; GENERAL INFORMATION:  
; APPLICANT: Bejanin, Stephanie  
; APPLICANT: Tanaka, Hiroaki  
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF  
; FILE REFERENCE: 91.US2.REG  
; CURRENT APPLICATION NUMBER: US/09/924,340  
; PRIOR FILING DATE: 2001-08-06  
; PRIOR APPLICATION NUMBER: US 60/305,456  
; PRIOR FILING DATE: 2001-07-13  
; PRIOR APPLICATION NUMBER: US 60/302,277  
; PRIOR FILING DATE: 2001-06-29  
; PRIOR APPLICATION NUMBER: US 60/298,698  
; PRIOR FILING DATE: 2001-06-15  
; PRIOR APPLICATION NUMBER: US 60/293,574  
; PRIOR FILING DATE: 2001-05-25  
; NUMBER OF SEQ ID NOS: 112  
; SOFTWARE: JPatent  
; SEQ ID NO 5  
; LENGTH: 438  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: 5'UTR  
; LOCATION: 1..83  
; NAME/KEY: CDS  
; LOCATION: 84..317  
; NAME/KEY: 3'UTR  
; LOCATION: 318..438  
; NAME/KEY: polyA\_signal  
; LOCATION: 397..402  
; NAME/KEY: polyA\_site  
; LOCATION: 423..438  
US-09-924-340-5

Query Match

Best Local Similarity 7.9%; Score 104.6; DB 9; Length 438;

Matches 113; Conservative 88.3%; Pred. No. 4.6e-17;

Mismatches 15; Indels 0; Gaps 0;







Db 2276 AAGAGAGGAGTACTTTCTAGTTAAATGAATGTAACCGTACTCAGGAAATGTTTTAA 2335  
 QY 997 TATATCCCTTAGTTCACAAAGTTAAATATCTTATTTCTGGATATGCTTTTATACCAAG 1056  
 Db 2336 AAGACTGCCAGACTCTTTTAAGGCTTTAACTATTTGTTATCTTCTGCTTTTATCTAACT 2395  
 QY 1057 AGCCTTATCAGCCAGTTCAGACACACCACTATACGCAACAACATACCAACAATACC 1116  
 Db 2396 ATAGGAATATAGATAATTTACTGAGAAAGATGAATTAATAATTAGTTTATTTGACATTTT 2455  
 QY 1117 AACATATACGTTTAAATATCATCTAGTAACCTGCAGGACATGATTTATGAGGCTTGATGG 1176  
 Db 2456 AAATAAATATGATTTTCTTATATGTATATAAATAAATAAACCCTGTTATATTTGATTTT 2515  
 QY 1177 CAATACGACTTCTACATCCATATCTCATCTTTTATACCATATACCACTACTACCACTT 1236  
 Db 2516 GTAAACAACATTTTAAAGTACAGCCTATATTTTAAATCTGTAGGCTAGATCAGAATT 2575  
 QY 1237 TTTGTNAGATCA 1248  
 Db 2576 AGTGCCATAGCA 2587

## RESULT 11

US-09-764-869-1744  
 ; Sequence 1744, Application US/09764869  
 ; Patent No. US20020061521A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rosen et al.  
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
 ; FILE REFERENCE: PC007  
 ; CURRENT APPLICATION NUMBER: US/09/764,869  
 ; CURRENT FILING DATE: 2001-01-17  
 ; Prior application data removed - refer to PALM or file wrapper  
 ; NUMBER OF SEQ ID NOS: 2442  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 1744  
 ; LENGTH: 7809  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-09-764-869-1744

Query Match 3.2%; Score 42.2; DB 10; Length 7809;  
 Best Local Similarity 45.8%; Pred. No. 2.1;  
 Matches 143; Conservative 0; Mismatches 169; Indels 0; Gaps 0;

QY 937 AATAATAAAACTCTGGTCTCCCTTAAGAAAAAACCCTTCCACCTTTACTGTGTCATT 996  
 Db 2277 AAGAGAGGAGTACTTTCTAGTTAAATGAATGTAACCGTACTCAGGAAATGTTTTAA 2336  
 QY 997 TATATCCCTTAGTTCACAAAGTTAAATATCTTATTTCTGGATATGCTTTTATACCAAG 1056  
 Db 2337 AAGACTGCCAGACTCTTTTAAGGCTTTAACTATTTGTTATCTTCTGCTTTTATCTAACT 2396  
 QY 1057 AGCCTTATCAGCCAGTTCAGACACACCACTATACGCAACAACATACCAACAATACC 1116  
 Db 2397 ATAGGAATATAGATAATTTACTGAGAAAGATGAATTAATAATTAGTTTATTTGACATTTT 2456  
 QY 1117 AACATATACGTTTAAATATCATCTAGTAACCTGCAGGACATGATTTATGAGGCTTGATGG 1176  
 Db 2457 AAATAAATATGATTTTCTTATATGTATATAAATAAATAAACCCTGTTATATTTGATTTT 2516  
 QY 1177 CAATACGACTTCTACATCCATATCTCATCTTTTATACCATATACCACTACTACCACTT 1236  
 Db 2517 GTAAACAACATTTTAAAGTACAGCCTATATTTTAAATCTGTAGGCTAGATCAGAATT 2576  
 QY 1237 TTTGTNAGATCA 1248  
 Db 2577 AGTGCCATAGCA 2588

## RESULT 12

US-09-764-869-1745

; Sequence 1745, Application US/09764869  
 ; Patent No. US20020061521A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rosen et al.  
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
 ; FILE REFERENCE: PC007  
 ; CURRENT APPLICATION NUMBER: US/09/764,869  
 ; CURRENT FILING DATE: 2001-01-17  
 ; Prior application data removed - refer to PALM or file wrapper  
 ; NUMBER OF SEQ ID NOS: 2442  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 1745  
 ; LENGTH: 7809  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-09-764-869-1745

Query Match 3.2%; Score 42.2; DB 10; Length 7809;  
 Best Local Similarity 45.8%; Pred. No. 2.1;  
 Matches 143; Conservative 0; Mismatches 169; Indels 0; Gaps 0;

QY 937 AATAATAAAACTCTGGTCTCCCTTAAGAAAAAACCCTTCCACCTTTACTGTGTCATT 996  
 Db 2276 AAGAGAGGAGTACTTTTCTAGTTAAATGAATGTAACCGTACTCAGGAAATGTTTTAA 2335  
 QY 997 TATATCCCTTAGTTCACAAAGTTAAATATCTTATTTCTGGATATGCTTTTATACCAAG 1056  
 Db 2336 AAGACTGCCAGACTCTTTTAAGGCTTTAACTATTTGTTATCTTCTGCTTTTATCTAACT 2395  
 QY 1057 AGCCTTATCAGCCAGTTCAGACACACCACTATACGCAACAACATACCAACAATACC 1116  
 Db 2396 ATAGGAATATAGATAATTTACTGAGAAAGATGAATTAATAATTAGTTTATTTGACATTTT 2455  
 QY 1117 AACATATACGTTTAAATATCATCTAGTAACCTGCAGGACATGATTTATGAGGCTTGATGG 1176  
 Db 2456 AAATAAATATGATTTTCTTATATGTATATAAATAAACCCTGTTATATTTGATTTT 2515  
 QY 1177 CAATACGACTTCTACATCCATATCTCATCTTTTATACCATATACCACTACTACCACTT 1236  
 Db 2516 GTAAACAACATTTTAAAGTACAGCCTTATATTTTAAATCTGTAGGCTAGATCAGAATT 2575  
 QY 1237 TTTGTNAGATCA 1248  
 Db 2576 AGTGCCATAGCA 2587

## RESULT 13

US-10-092-154-1685  
 ; Sequence 1685, Application US/10092154  
 ; Publication No. US20030054375A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rosen et al.  
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
 ; FILE REFERENCE: PC009C1  
 ; CURRENT APPLICATION NUMBER: US/10/092,154  
 ; CURRENT FILING DATE: 2002-03-07  
 ; NUMBER OF SEQ ID NOS: 2003  
 ; Prior Application removed - See File Wrapper or Palm  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 1685  
 ; LENGTH: 556  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-10-092-154-1685

Query Match 3.1%; Score 40.6; DB 9; Length 556;  
 Best Local Similarity 55.2%; Pred. No. 1.3;  
 Matches 79; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 575 TTTTGTACTTTATCTGAATCCAGATGCTGCTCCCTCGCAGGTTGTTTCCCTTTCTTA 634  
 Db 88 TTTAAAAAATTTTTTTTAAATGCACCTCTTCTCCCGTGTGGATATATCATCTACT 147

Search completed: April 14, 2003, 23:05:38  
Job time : 608 secs

Journal of Management Inquiry 22(4) 403-420

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 14, 2003, 21:02:56 ; Search time 1828 Seconds  
(without alignments)  
11792.228 Million cell updates/sec

Title: US-09-830-244B-2  
Perfect score: 1331  
Sequence: 1 ctatgttttttagaatcaag.....gccttttccaatgctacttg 1331

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

#### Database :

EST.\*  
1: em\_estba.\*  
2: em\_esthum.\*  
3: em\_estlin.\*  
4: em\_estnu.\*  
5: em\_estov.\*  
6: em\_estpl.\*  
7: em\_estro.\*  
8: em\_htc.\*  
9: gb\_est1.\*  
10: gb\_est2.\*  
11: gb\_htc.\*  
12: gb\_est3.\*  
13: gb\_est4.\*  
14: gb\_est5.\*  
15: em\_estfun.\*  
16: em\_estom.\*  
17: gb\_gss.\*  
18: em\_gss\_hum.\*  
19: em\_gss\_inv.\*  
20: em\_gss\_pln.\*  
21: em\_gss\_vrt.\*  
22: em\_gss\_fun.\*  
23: em\_gss\_mam.\*  
24: em\_gss\_mus.\*  
25: em\_gss\_other.\*  
26: em\_gss\_pro.\*  
27: em\_gss\_rod.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	ID	Description
C 1	490	36.8	492	9	AI224097
C 2	480	36.1	518	9	AI694371
C 3	478.4	35.9	523	10	AW269813
C 4	405	30.4	888	12	BG121534
C 5	399	30.0	427	12	BF831772
C 6	398.4	29.9	417	12	BE767736

C 7	380.4	28.6	394	12	BG011356
C 8	369.4	27.8	403	12	BF831960
C 9	353	26.5	408	12	BG011354
C 10	352.8	26.5	392	12	BE929791
C 11	345.6	26.0	402	12	BG011351
C 12	344.8	25.9	395	12	BF996088
C 13	335.2	25.2	951	12	BG168510
C 14	333	25.0	396	12	BF832034
C 15	330.4	24.8	376	9	AA488304
C 16	332	24.2	386	14	BQ359015
C 17	309.2	23.2	352	14	BQ359015
C 18	293.2	22.0	324	12	BF986192
C 19	280.8	21.1	417	14	BQ359018
C 20	280	21.0	333	12	BF832937
C 21	266.6	20.0	295	10	AW135014
C 22	262.4	19.7	279	12	BF195348
C 23	252.6	19.0	365	10	BE089961
C 24	249.8	18.8	517	12	BG188342
C 25	248.8	18.7	470	12	BG200817
C 26	246.2	18.5	404	9	AA376713
C 27	245.2	18.4	405	9	AA376719
C 28	242.4	18.2	519	10	AW951480
C 29	239	18.0	250	9	AI630950
C 30	238.4	17.9	410	14	T29650
C 31	236	17.7	375	9	AA376681
C 32	228.6	17.2	335	9	AA376714
C 33	222	16.7	257	12	BF832847
C 34	211	15.9	457	9	AA376682
C 35	203.4	15.3	306	10	BE163886
C 36	197.8	14.9	426	9	AA376680
C 37	181.4	13.6	302	9	AA376694
C 38	179.2	13.5	341	9	AA376695
C 39	175.6	13.2	329	14	BQ309679
C 40	169.6	12.7	321	12	BG205925
C 41	163	12.2	258	12	BF834202
C 42	161.4	12.1	314	10	AW409123
C 43	161	12.1	209	12	BF835534
C 44	155.6	11.7	173	12	BF088945
C 45	155	11.6	172	12	BF944749

#### ALIGNMENTS

RESULT 1  
AI224097/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
FEATURES  
source

AI224097  
gi35a03.x1 Soares\_NHMPu\_S1 Homo sapiens cDNA clone IMAGE:1858444  
3', mRNA sequence.

AI224097  
AI224097.1 GI:3806810  
EST.  
human.

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 492)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP).  
Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgaps-r@mail.nih.gov](mailto:cgaps-r@mail.nih.gov)  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Insert Length: 876 Std Error: 0.00  
Seq primer: -40UP from Gibco  
High quality sequence stop: 456.  
Location/Qualifiers  
1..492  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1858444"



/clone\_lib="Soares\_NhMPu\_S1"  
 /tissue\_type="Pooled human melanocyte, fetal heart, and  
 pregnant uterus"  
 /lab\_host="DH10B"  
 /note="Organ: mixed (see below); Vector: pT7T3D-Pac  
 (Pharmacia) with a modified polylinker; Site: 1: Not I;  
 Site 2: Eco RI; Equal amounts of plasmid DNA from three  
 normalized libraries (melanocyte 2NDHM, pregnant uterus  
 NbHPu, and fetal heart NbHH19) were mixed, and ss circles  
 were made in vitro. Following HAP purification, this DNA  
 was used as tracer in a subtractive hybridization  
 reaction. The driver was PCR-amplified cDNAs from pools of  
 5,000 clones made from the same 3 libraries. The pools  
 consisted of I.M.A.G.E. clones 260232-265223,  
 340488-345479, and 484488-489479."  
 BASE COUNT 167 a 103 c 84 g 136 t  
 ORIGIN

Query Match 36.8%; Score 490; DB 9; Length 492;  
 Best Local Similarity 99.6%; Pred. No. 8.1e-105;  
 Matches 490; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 9 TTGAATCAAGATGAACCGGTAGCTCTCATGTACCAAGCGTGAATTTACAGTGT 68  
 Db TTAGAATCANAGATGAACCGGTAGCTCTCATGTACCAAGCGTGAATTTACAGTGT 433  
 QY 69 TACAAATGCTCGGAATTTTCACCTGCCATAGGGAATGTTAAGTTACTTGGCTGGAAATTT 128  
 Db TACAAATGCTCGGAATTTTCACCTGCCATAGGGAATGTTAAGTTACTTGGCTGGAAATTT 433  
 QY 129 ATCAGACTGTGAGTAAACAAAGTTGAAGTTTACAGATGAGGGGAATATTGAGGCCCT 188  
 Db ATCAGACTGTGAGTAAACAAAGTTGAAGTTTACAGATGAGGGGAATATTGAGGCCCT 313  
 QY 189 AAGCTAAACAAATATCATCATGATGCTGCTAATGTGCTCCCGAGGCTAAT 248  
 Db AAGCTAAACAAATATCATCATGATGCTGCTAATGTGCTCCCGAGGCTAAT 253  
 QY 249 TTGGGAACAGTTTTTCTGATGCTTTTATAGAGCATTTGATTTCTTCTTTTGGG 368  
 Db TTGGGAACAGTTTTTCTGATGCTTTTATAGAGCATTTGATTTCTTCTTTTGGG 193  
 QY 309 CTGCTGCCATTTAGCATTTTTCAGATCTTTTATAGAGCATTTGATTTCTTCTTTTGGG 368  
 Db CTGCTGCCATTTAGCATTTTTCAGATCTTTTATAGAGCATTTGATTTCTTCTTTTGGG 133  
 QY 369 GAAATGTCCTATTTAGCATTTTTCAGATCTTTTATAGAGCATTTGATTTCTTCTTTTGGG 428  
 Db GAAATGTCCTATTTAGCATTTTTCAGATCTTTTATAGAGCATTTGATTTCTTCTTTTGGG 73  
 QY 429 CCGTTATTTGTTGATACAGCATTTGATTTTACAGTAACTGTTTACAGTAACTGTTTACAGTAACT 488  
 Db CCGTTATTTGTTGATACAGCATTTGATTTTACAGTAACTGTTTACAGTAACTGTTTACAGTAACT 13  
 QY 489 TGCTGCACCTCT 500  
 Db TGCTGCACCTCT 1

RESULT 2  
 A1694371/c  
 LOCUS  
 DEFINITION wd83a04.x1 NCI\_CGAP\_Lu24 Homo sapiens cdna clone IMAGE:2338158 3',  
 mRNA sequence.  
 ACCESSION A1694371  
 VERSION A1694371.1 GI:4971711  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 518)  
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

# TITLE JOURNAL COMMENT

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
 Emmert-Buck, M.D., Ph.D.  
 cDNA Library Preparation: M. Bento Soares, Ph.D.  
 cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)  
 Insert Length: 710 Std Error: 0.00  
 Seq primer: -400P from GIBCO  
 High quality sequence stop: 454.  
 Location/Qualifiers  
 1..518  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2338158"  
 /clone\_lib="NCI\_CGAP\_Lu24"  
 /tissue\_type="carcinoid"  
 /lab\_host="DH10B"  
 /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a  
 modified polylinker; plasmid DNA from the normalized  
 library NCI\_CGAP\_Lu5 was prepared, and ss circles were  
 made in vitro. Following HAP purification, this DNA was  
 used as tracer in a subtractive hybridization reaction.  
 The driver was PCR-amplified cDNAs from a pool of 5,000  
 clones made from the same library (cloneIDs  
 1414920-1417991 and 1520904-1522439). Subtraction by Bento  
 Soares and M. Fatima Bonaldo."  
 BASE COUNT 171 a 107 c 83 g 157 t  
 ORIGIN

Query Match 36.1%; Score 480; DB 9; Length 518;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-102;  
 Matches 480; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CTATGTTTTAGAAATCAACCGGTAAAGCTGCTCATGTACCAACCGTGAATTT 60  
 Db CTATGTTTTAGAAATCAACCGGTAAAGCTGCTCATGTACCAACCGTGAATTT 430  
 QY 61 ACAGTGTTTACAAATGCTGGAAATTTGCACCTGCCATAGGGAATGTTAAGTTTAC 120  
 Db ACAGTGTTTACAAATGCTGGAAATTTGCACCTGCCATAGGGAATGTTAAGTTTAC 370  
 QY 121 TGGAAATTTATCAGACTGTGAGTAAACAAAGTTGAAGTTTACGAGATGAGGGGAATTTG 180  
 Db TGGAAATTTATCAGACTGTGAGTAAACAAAGTTGAAGTTTACGAGATGAGGGGAATTTG 310  
 QY 181 AGGCCCTTAAGGCTAAACAAATAATCACTATCTGAGATAGTGGCTAATGTGGCTCCCA 240  
 Db AGGCCCTTAAGGCTAAACAAATAATCACTATCTGAGATAGTGGCTAATGTGGCTCCCA 250  
 QY 241 GGCTTAATTTGGGAACAGTTTTTCTCGATGCTTTTGAGAAGTACTTTCTTTTGACAGAA 300  
 Db GGCTTAATTTGGGAACAGTTTTTCTCGATGCTTTTGAGAAGTACTTTCTTTTGACAGAA 190  
 QY 301 TTTTCATCTGCTGGCAATGCTATATCTCCCTTTATAGGAGCCATTTGATTTCTTTCC 360  
 Db TTTTCATCTGCTGGCAATGCTATATCTCCCTTTATAGGAGCCATTTGATTTCTTTCC 130  
 QY 361 TTTTGTGGGAATGTCCTCAATTTAGCAATTTTCAGATCTTTTGTGTCACATTAATGCCATTTAT 420  
 Db TTTTGTGGGAATGTCCTCAATTTAGCAATTTTCAGATCTTTTGTGTCACATTAATGCCATTTAT 70  
 QY 421 TGGTAATCCGCTTATTTGGTGAATACAGCATAGTAAATAAAGTCTGTACAGTAAATCTACA 480  
 Db TGGTAATCCGCTTATTTGGTGAATACAGCATAGTAAATAAAGTCTGTACAGTAAATCTACA 10

```

RESULT 3
AW269813/c
LOCUS
DEFINITION
  xv45f02.x1 Soares_NFL_T_GBC_S1 Homo sapiens linear EST 03-JAN-2000
  IMAGE:2816091 3', mRNA sequence.
ACCESSION
  AW269813
VERSION
  AW269813.1 GI:6656843
KEYWORDS
  EST.
SOURCE
  human.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
  NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
  Tumor Gene Index
  Unpublished (1997)
  Contact: Robert Strausberg, Ph.D.
  Email: cgapbs-remail.nih.gov
  This clone is available royalty-free through LLNL; contact the
  IMAGE Consortium (Info@image.llnl.gov) for further information.
  Seq primer: -40UP from Gibco
  High quality sequence stop: 462.
FEATURES
  source
    1..523
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="IMAGE:2816091"
    /lab_host="Soares_NFL_T_GBC_S1"
    /lab_host="DH10B"
    /notes="Organ: pooled; Vector: pMT73D-Pac (Pharmacia) with
    a modified polylinker; Site_1: Not I; Site_2: Eco RI;
    Equal amounts of plasmid DNA from three normalized
    libraries (fetal lung MbHL19w, testis NHT, and B-cell
    NCI-CGAP_GCB1) were mixed, and ss circles were made in
    vitro. Following HAP purification, this DNA was used as
    tracer in a subtractive hybridization reaction. The driver
    was PCR-amplified cDNAs from pools of 5,000 clones made
    from the same 3 libraries. The pools consisted of
    I.M.A.G.E. clones 297480-302087, 682632-687239,
    726408-728711, and 729096-731399. Subtraction by Bento
    Soares and M. Fatima Bonaldo. "
    177 a 109 c 85 g 152 t
BASE COUNT
  177 a 109 c 85 g 152 t
ORIGIN
  Query Match 35.9%; Score 478.4; DB 10; Length 523;
  Best Local Similarity 99.8%; Pred. No. 4.3e-102;
  Matches 479; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
  QY 1 CTATGTTTTAGAAATCAAGATGAACCGGTAAAGCTGTCATGTCACCAACCGTGAATTT 60
  DB 481 CTATGTTTTAGAAATCAAGATGAACCGGTAAAGCTGTCATGTCACCAACCGTGAATTT 422
  QY 61 ACAGTCTTTACAAATGCTCGGAATTTGCACTGCCATAGGGAATGTTAAGGTTACTTGGC 120
  DB 421 ACAGTCTTTACAAATGCTCGGAATTTGCACTGCCATAGGGAATGTTAAGGTTACTTGGC 362
  QY 121 TGGAAATTTATCAGACTTGTGAGTAAACAGTTGAAGTTAGCAGATGAGGGGAATATTG 180
  DB 361 TGGAAATTTATCAGACTTGTGAGTAAACAGTTGAAGTTAGCAGATGAGGGGAATATTG 302
  QY 181 AGGCCCTTAAGGCTAAACAAATAATCAGTATCTGAGATAGTGGCTAATGTCGCTCCCA 240
  DB 301 AGGCCCTTAAGGCTAAACAAATAATCAGTATCTGAGATAGTGGCTAATGTCGCTCCCA 242
  QY 241 GGCCTAAATTTGGGAACAGTTTTCCTGATCTGCTTTGAGAGTACTTCTTTTGACAGAAA 300
  DB 241 GGCCTAAATTTGGGAACAGTTTTCCTGATGCTTTGAGAGTACTTCTTTTGACAGAAA 182
  QY 301 TTTTCATCTCTGCTGCGCATTCGTATATCTCCCTTTATAGGAGCCATTTGATTTCTTCC 360
  DB 181 TTTTCATCTCTGCTGCGCATTCGTATATCTCCCTTTATAGGAGCCATTTGATTTCTTCC 122

```

```

QY 361 TTTTGTGGAAATGTCCTCCATAGCATTTTCAGATCTTTTCATGTGCACTAATGCCATTAT 420
DB 121 TTTTGTGGAAATGTCCTCCATAGCATTTTCAGATCTTTTCATGTGCACTAATGCCATTAT 62
QY 421 TGGTAATGCCGTTATTGTTGGTGAATACAGCATAGTAAATAAATCTTTACAGTAATCTACA 480
DB 61 TGGTAATGCCGTTATTGTTGGTGAATACAGCATAGTAAATAAATCTTTACAGTAATCTACA 2

RESULT 4
BG121534
LOCUS
DEFINITION
  602352872F1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:4451201 5',
  mRNA sequence.
ACCESSION
  BG121534
VERSION
  BG121534.1 GI:12615043
KEYWORDS
  EST.
SOURCE
  human.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
  NIH-MGC http://mgc.nci.nih.gov/.
  National Institutes of Health, Mammalian Gene Collection (MGC)
  Unpublished (1999)
  Contact: Robert Strausberg, Ph.D.
  Email: cgapbs-remail.nih.gov
  Tissue Procurement: ATCC
  cDNA Library Preparation: Life Technologies, Inc.
  DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
  Clone distribution: MGC clone distribution information can be
  found through the I.M.A.G.E. Consortium/LLNL at:
  http://image.llnl.gov
  Plate: LLAM10237 row: p column: 18
  High quality sequence start: 2
  High quality sequence stop: 680.
FEATURES
  source
    1..888
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="IMAGE:4451201"
    /lab_host="NIH_MGC_90"
    /tissue_type="adenocarcinoma, cell line"
    /notes="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
    Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
    Average insert size 1.7 kb. Library enriched for
    full-length clones and constructed by Life Technologies.
    Note: this is a NIH_MGC Library."
    240 a 161 c 198 g 289 t
BASE COUNT
  240 a 161 c 198 g 289 t
ORIGIN
  Query Match 30.4%; Score 405; DB 12; Length 888;
  Best Local Similarity 89.8%; Pred. No. 7.9e-85;
  Matches 468; Conservative 0; Mismatches 50; Indels 3; Gaps 3;
  QY 1 CTATGTTTTAGAAATCAAGATGAACCGGTAAAGCTGTCATGTCACCAACCGTGAATTT 60
  DB 284 CTATGTTTTAGAAATCAAGATGAACCGGTAAAGCTGTCATGTCACCAACCGTGAATTT 343
  QY 61 ACAGTCTTTACAAATGCTCGGAATTTGCACTGCCATAGGGAATGTTAAGGTTACTTGGC 120
  DB 344 ACAGTCTTTACAAATGCTCGGAATTTGCACTGCCATAGGGAATGTTAAGGTTACTTGGC 403
  QY 121 TGGAAATTTATCAGACTTGTGAGTAAACAGTTGAAGTTAGCAGATGAGGGGAATATTG 180
  DB 404 TGGAAATTTATCAGACTTGTGAGTAAACAGTTGAAGTTAGCAGATGAGGGGAATATTG 463
  QY 181 AGGCCCTTAAGGCTAAACAAATAATCAGTATCTGAGATAGTGGCTAATGTCGCTCCCA 239
  DB 464 AGGCCCTTAAGGCTAAACAAATAATCAGTATCTGAGATAGTGGCTAATGTCGCTCCCA 523

```



/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="GN0042"  
/dev\_stages="Adult"

/note="Organ: placenta.normal; Vector: puc18; Site\_1: SmaI products derived from ORESIES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT  
ORIGIN

143 a 92 c 71 g 111 t

Query Match 29.9%; Score 398.4; DB 12; Length 417;  
Best Local Similarity 99.8%; Pred. No. 2.9e-83;  
Matches 399; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 42 TGTACCAACCTGAAATTTACAGTGTTCACAAATGCTCTGGAATTTTGCACCTGCCATAGGG 101  
|||||  
Db 417 TGTACCAACCTGAAATTTACAGTGTTCACAAATGCTCTGGAATTTTGCACCTGCCATAGGG 358  
QY 102 AATGTTAAGTTACTTGGCTGGAATTTATCAGACTTGTGAGTAACAAGTTGAAGTTAG 161  
|||||  
Db 357 AATGTTAAGTTACTTGGCTGGAATTTATCAGACTTGTGAGTAACAAGTTGAAGTTAG 298  
QY 162 CAGATGAGGGGAATATTGAGGCCCTTAAGCTTAACAAAATATCAGTATCTGAGATAG 221  
|||||  
Db 297 CAGATGAGGGGAATATTGAGGCCCTTAAGCTTAACAAAATATCAGTATCTGAGATAG 238  
QY 222 TGGCTAATGTGGCTCCCGAGCCCTAAATTTGGGAACAGTTTTCCTGATTCCTTTATAGG 281  
|||||  
Db 237 TGGCTAATGTGGCTCCCGAGCCCTAAATTTGGGAACAGTTTTCCTGATTCCTTTGAGAAG 178  
QY 282 TACTTTCTTTGACAGAAATTTTCATCTGCTTGCCTTGCCTATATCTCCCTTTATAGG 341  
|||||  
Db 177 TACTTTCTTTGACAGAAATTTTCATCTGCTTGCCTTGCCTATATCTCCCTTTATAGG 118  
QY 342 AGCCATTGGATTCTTTCCTTTTGTGGGAATGTCCTCCATAGCATTTTTCAGATCTTTTGA 401  
|||||  
Db 117 AGCCATTGGATTCTTTCCTTTTGTGGGAATGTCCTCCATAGCATTTTTCAGATCTTTTGA 58  
QY 402 TGTGCACTAATGCCATTATTGGTAATGCCGTTATTGGTGA 441  
|||||  
Db 57 TGAGCACTAATGCCATTATTGGTAATGCCGTTATTGGTGA 18

RESULT 7

LOCUS BG011356/c

DEFINITION RC1-GN0268-091200-011-g04 GN0268 Homo sapiens linear EST 24-JAN-2001

ACCESSION BG011356

VERSION BG011356.1 GI:12459473

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 394)  
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

TITLE

JOURNAL

MEDLINE

COMMENT

RESULT 8

LOCUS BF831960

DEFINITION

PM3-HT0909-181000-010-g05 HT0909 Homo sapiens linear EST 13-JAN-2001

ACCESSION BF831960

VERSION BF831960.1 GI:12180211

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 403)

REFERENCE

AUTHORS

Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,

Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gechtml2.pl?cl=RC1&t2=RC1-GN0268-091200-011-g04&t3=2000-12-09&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 17  
High quality sequence stop: 394.

FEATURES  
source

1..394  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="GN0268"  
/dev\_stages="Adult"  
/note="Organ: placenta.normal; Vector: puc18; Site\_1: SmaI products derived from ORESIES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."  
BASE COUNT 137 a 86 c 67 g 104 t

Query Match 28.6%; Score 380.4; DB 12; Length 394;  
Best Local Similarity 99.2%; Pred. No. 4.9e-79;  
Matches 393; Conservative 0; Mismatches 1; Indels 2; Gaps 1;

QY 42 TGTACCAACCTGAAATTTACAGTGTTCACAAATGCTCTGGAATTTTGCACCTGCCATAGGG 101  
|||||  
Db 394 TGTACCAACCTGAAATTTACAGTGTTCACAAATGCTCTGGAATTTTGCACCTGCCATAGGG 335  
QY 102 AATGTTAAGTTACTTGGCTGGAATTTATCAGACTTGTGAGTAACAAGTTGAAGTTAG 161  
|||||  
Db 334 AATGTTAAGTTACTTGGCTGGAATTTATCAGACTTGTGAGTAACAAGTTGAAGTTAG 275  
QY 162 CAGATGAGGGGAATATTGAGGCCCTTAAGCTTAACAAAATATCAGTATCTGAGATAG 221  
|||||  
Db 274 CAGATGAGGGGAATATTGAGGCCCTTAAGCTTAACAAAATATCAGTATCTGAGATAG 215  
QY 222 TGGCTAATGTGGCTCCCGAGCCCTAAATTTGGGAACAGTTTTCCTGATTCCTTTATAGG 281  
|||||  
Db 214 TGGCTAATGTGGCTCCCGAGCCCTAAATTTGGGAACAGTTTTCCTGATTCCTTTGAGAAG 155  
QY 282 TACTTTCTTTGACAGAAATTTTCATCTGCTTGCCTTGCCTATATCTCCCTTTATAGG 341  
|||||  
Db 154 TACTTTCTTTGACAGAAATTTTCATCTGCTTGCCTTGCCTATATCTCCCTTTATAGG 95  
QY 342 AGCCATTGGATTCTTTCCTTTTGTGGGAATGTCCTCCATAGCATTTTTCAGATCTTTTGA 401  
|||||  
Db 94 AGCCATTGGATTCTTTCCTTTTGTGGGAATGTCCTCCATAGCATTTTTCAGATCTTTTGA 37  
QY 402 TGTGCACTAATGCCATTATTGGTAATGCCGTTATTGG 437  
|||||  
Db 36 TGTGCACTAATGCCATTATTGGTAATGCCGTTATTGG 1

RESULT 8

LOCUS BF831960

DEFINITION

PM3-HT0909-181000-010-g05 HT0909 Homo sapiens linear EST 13-JAN-2001

ACCESSION BF831960

VERSION BF831960.1 GI:12180211

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 403)

REFERENCE

AUTHORS

Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,

TITLE	Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
JOURNAL MEDLINE	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
COMMENT	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000) Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL <a href="http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM3&amp;t2=PM3-HF0909-181000-010-q05&amp;t3=2000-10-18&amp;t4=1">http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM3&amp;t2=PM3-HF0909-181000-010-q05&amp;t3=2000-10-18&amp;t4=1</a> Seq primer: puc 18 forward High quality sequence start: 39 High quality sequence stop: 403.

[illegible]

BASE COUNT	105 a	68 c	88 g	142 t	
ORIGIN	low stringency conditions."				
Query Match	27.8%; Score 369.4; DB 12; Length 403;				
Best Local Similarity	99.7%; Pred. No. 1.9e-76;				
Matches 370; Conservative	0; Mismatches 1; Indels 0; Gaps 0;				
Qy	68	TTACAAATGCTCGGAATTTTCACATGCCATAGGGAATGTAAGGTACTTCGCTGGGAATT	127		
Db	33	TTACAAATGCTCGGAATTTTCACATGCCATAGGGAATGTAAGGTACTTCGCTGGGAATT	92		
Qy	128	TATCAGACTTCTGAGTAAACAAGTTGAAGTTTACAGATGAGGGGAATATTGAGGCCCC	187		
Db	93	TATCAGACTTCTGAGTAAACAAGTTGAAGTTTACAGATGAGGGGAATATTGAGGCCCC	152		
Qy	188	TAAAGCTAAACAAAATATACAGTATCTGAGATCTGCGCTAATGTGGCTCCCGAGGCCATA	247		
Db	153	TAAAGCTAAACAAAATATACAGTATCTGAGATCTGCGCTAATGTGGCTCCCGAGGCCATA	212		
Qy	248	TTTGGGAACAGTTTTTCCCTGATTCGTTTGAGAAGTACTTCTTTTGACAGAAATTTTCAT	307		
Db	213	TTTGGGAACAGTTTTTCCCTGATTCGTTTGAGAAGTACTTCTTTTGACAGAAATTTTCAT	272		
Qy	308	TCGCTTGCCATTGCTATATTCGCCCTTTATAGAGCCATTGGATTTCTTCCTTTTGTG	367		
Db	273	TCGCTTGCCATTGCTATATTCGCCCTTTATAGAGCCATTGGATTTCTTCCTTTTGTG	332		
Qy	368	GGAATATGCCATTAGCATTTTCAGATCTTTGATGTGCACTAATGCCATTATGGTAAT	427		
Db	333	GGAATATGCCATTAGCATTTTCAGATCTTTGATGTGCACTAATGCCATTATGGTAAT	392		
Qy	428	GCGGTTATGG 438			
Db	393	GCGGTTATGG 403			

RESULT 9	
BG011354/c	
LOCUS	
DEFINITION	
ACCESSION	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	

FEATURES  
SOURCE

BASE COUNT	142 a	90 c	71 g	105 t	performed under low stringency conditions."			
ORIGIN								
Query Match	26.5%	Score 353;	DB 12;	Length 408;				
Best Local Similarity	96.88;	Pred. No. 1.4e-72;						
Matches 392;	Conservative	0;	Mismatches 10;	Indels 3;	Gaps 3;			
QY 42	TGTACCAACGCGAAATTTACAGTGTTTACAAATGCTGGAAATTTGCACATGCCATAGGG	101						
Db	TGTACCAACGCGAAATTTACAGTGTTTACCAATGCTGCAATTTTGCACATGCCATAGGG	349						
QY 102	AATGTTAAAGCTTACTTGGCTGGGAATTTATCAGACTTGTGAGTAAACAAGTTGAAGTTTAG	161						
Db	AATGTTAAAGCTTACTTGGCTGGGAATTTATCAGACTTGTGAGTATACAAGTTGAAGTTTAG	289						
QY 162	CAGATCAGGGGGGAATATTGAGGCCCCCTAAGCGTAAACAAAATAATCAGTATCTCGAGATAG	221						
Db	CAGATCAGGGGGGAATATTGAGGCCCCCTAAGCGTAAACAAAATAATCAGTATCTCGAGATAG	229						
QY 222	TGCGCTAATGTGGCTCCCGCAGGCCCTAATTTGGGGAACAGTTTTTCTCGATTCGCTTTGAGAAG	281						
Db	TGCGCTAATGTGGCTCCCGCAGGCCCTAATTTGGGGAACAGTTTTTCTCGATTCGCTTTGAGAAG	169						





Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT	139 a	92 c	71 g	100 t	performed under low stringency conditions."
ORIGIN					
Query Match	26.0%;	Score 345.6;	DB 12;	Length 402;	
Best Local Similarity	94.8%;	Pred. No. 7.6e-71;			
Matches 368;	Conservative 0;	Mismatches 19;	Indels 1;	Gaps 1;	
QY	42	TGTACCAACGCGAAATTTACAGATGTTTACAAATGCTCGAAATTTTGCACTGCCATAGG	101		
Db	402	TGTACCGAGCGTGAATTTACAGTGTTCAGATGCTCGAAATTTTGCACTGCCATAGG	343		
QY	102	AATGTTAAGGTTACTTGCTCGAATTTATCAGACTTGTGTAGTAAACAAGTTGAAGTTTAG	161		
Db	342	AATGTTAAGGTTACTTGCTCGAATTTATCAGACTTGTGTAGTAAACAAGTTGAAGTTTAG	283		
QY	162	CAGATGAGGGGGAATATTGAGGCCCTTAAGGCTAAACAANAATCAGTATCTGAGATAG	221		
Db	282	CAGATGAGGGGGAATATTGAGGCCCTTAAGGCTAAACAANAATCAGTATCTGAGATAG	223		
QY	222	TGGCTAATGTGGCTCCCCAGGCTTAATTTGGGAAACAGTTTTTCCCTGATTCGTTGAGAG	281		
Db	222	TGGCTAATGTGGCTCCCCAGGCTTAATTTGGGAAACAGTTTTTCCCTGATTCGTTGAGAG	163		
QY	282	TACTTCTCTTTTGACAGAAATTTTCATTCGTTGCCATGCTATATTTCTCCCTTTATAGG	341		
Db	162	TACTTCTCTTTTGACAGAAATTTTCATTCGTTGCCATGCTATATTTCTCCCTTTATAGG	103		
QY	342	AGCCATTGGATTTCTTTCCCTTTTGTGGGAATGTCCTCATTCAGATTTTCAGATCTTTTGA	401		
Db	102	AGCCATTGGATTTCTTTCCCTTTTGTGGGAATGTCCTCATTCAGATTTTCAGATCTTTTGA	43		
QY	402	TGTGCACTAATGCCATTTATTGGTAAATGC	429		
Db	42	TGTGCTCTAATGCCATTTATTGGTAAATGC	16		

RESULT 12	BF996088	BF996088	395 bp	linear	EST 23-JAN-2007
LOCUS					
DEFINITION		RC3-GN0042-061100-019-e07	GN0042	Homo sapiens	cdNA, mRNA sequence.
ACCESSION		BF996088			
VERSION		BF996088.1	GI:12402411		
KEYWORDS		EST.			
SOURCE		human.			



Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Plate: LLAM10228 row: n column: 16  
High quality sequence stop: 675.  
Location/Qualifiers  
1. .951

FEATURES  
source

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4447695"  
/clone\_lib="NIH\_MGC\_89"  
/tissue\_type="hypertrophied, cell line"  
/lab\_host="PH108 (phage-resistant)"  
/note="Organ: kidney; Vector: pCMV-SPORT6; Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.3 Kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH\_MGC Library."  
Location/Qualifiers  
281 a 221 c 196 g 253 t

BASE COUNT  
ORIGIN

Query Match 25.2%; Score 335.2; DB 12; Length 951;  
Best Local Similarity 95.8%; Pred. No. 2.1e-68;  
Matches 366; Conservative 0; Mismatches 13; Indels 3; Gaps 2;

QY 709 GACAGAGCTCTGCTCAAAAAGGACATTTATCATATTAACATCTTATTAGACCC 768  
DB 76 GACAGAGCTCTGCTCAAAAAGGACATTTATCATATTAACATCTTATTAGACCC 135  
QY 769 TAATTTCTTATCTGAAGGACATCTTTTAAACAGTTAAGTACGATGCTCAACAG 828  
DB 136 TAATTTCTTATCTGAAGGACATCTTTTAAACAGTTAAGTACGATGCTCAACAG 193  
QY 829 ACAATATTTCTGATGATAGTATCCCTGTCACAGTAGCAATGTGGTTTCATAAAGTG 888  
DB 194 ACAATATTTCTGATGATAGTATCCCTGTCACAGTAGCAATGTGGTTTCATAAAGTG 253  
QY 889 GGAAGAAACACATTTTAAAGTAACTTTTGGGACAGTATGAGTAATAATAAACT 948  
DB 254 GGAAGAAACACATTTTAAAGTAACTTTTGGGACAGTATGAGTAATAATAAACT 312  
QY 949 CTGGTCTCCCTTAAGAAAAAACCCTCCACCTTTACTGTGTCATTTATATCCCTTA 1008  
DB 313 CTGGTCTCCCTTAAGAAAAAACCCTCCACCTTTACTGTGTCATTTATATCCCTTA 372  
QY 1009 GTTCAAAGTTAATATCTTATTTCTGGATATGCTTTTATACCAAGAGCCTTATCAGC 1068  
DB 373 GTTCAAAGTTAATATCTTATTTCTGGATATGCTTTTATACCAAGAGCCTTATCAGC 432  
QY 1069 CAGTTCCAGAACACCATATA 1090  
DB 433 CTTGTACTACAGTATCTTTA 454

RESULT 14  
BF832034  
LOCUS

PF832034 396 bp mRNA linear EST 13-JAN-2001  
DEFINITION PF832034  
VERSION PF832034.1 GI:12180354  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 396)  
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunschein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed

TITLE

JOURNAL  
MEDLINE  
COMMENT

sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
20202663  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (<http://www.ludwig.org.br/scripts/gethtml2.pl?cl=PM3&t2=PM3-HT0909-181000-011-h06&t3=2000-10-18&t4=1>)  
Seq primer: puc 18 forward  
High quality sequence stop: 87.  
Location/Qualifiers  
1. .396

FEATURES  
source

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="HT0909"  
/dev\_stage="Adult"  
/note="Organ: head\_neck; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."  
Location/Qualifiers  
104 a 57 c 96 g 139 t

BASE COUNT  
ORIGIN

Query Match 25.0%; Score 333; DB 12; Length 396;  
Best Local Similarity 96.2%; Pred. No. 7e-68;  
Matches 384; Conservative 0; Mismatches 10; Indels 5; Gaps 4;

QY 42 TGTACCAACGTTAAATTTACAGTCTTTACAAATCTCTGGAATTTTGCACTGCCATAGG 101  
DB 1 TGTACCAACGTTAAATTTACAGTCTTTACAAATCTCTGGAATTTTGCACTGCCATAGG 60  
QY 102 AATGTTAAGTTACTTGGCTGGAATTTATCAGACT--TGTGAGTAACAAGTTGAAGTTT 159  
DB 61 AATGTTAAGG-TACTTGGTGGAAATTTATTAGAAATGTGTGAGTAACAAGTTGAAGTTT 119  
QY 160 AGCAGATCAGGGGGAATATTGAGGCCCTTAAGGCTTAACAAATATCAGTATCTCAGAT 219  
DB 120 AGCAGATCAGGGGGAATATTGAGGCCCTTAAGGCTTAACAAATATCAGTATCTCAGAT 178  
QY 220 AGTGCTAATGTGGCTCCCGAGGCTAATTTGGGAACAGTTTTCCTGATTCCTTTGAGA 279  
DB 179 AGTGCTAATGTGGCTCCCGAGGCTAATTTGGGAACAGTTTTCCTGATTCCTTTGAGA 238  
QY 280 AGTACTTCTTTTACACAGAAATTTTCATTCGTTGCCATTCGCTATATTCCTCTTATA 339  
DB 239 AGTACTTCTTTTACACAGAAATTTTCATTCG-TGTGGCAATTCGCTATATTCCTCTTATA 297  
QY 340 GGAGCCATTGGATTTCTTTCTCTTTTGTGGGAAATGTCCTCATTTAGCATTTTCAGATCTTT 399  
DB 298 GGAGCCATTGGATTTCTTTCTCTTTTGTGGGAAATGTCCTCATTTAGCATTTTCAGATCTTT 357  
QY 400 GATGTGCACTAATGCCATTTATTGGTAATGCCGTTATTGG 438  
DB 358 GATGTGCACTAATGCCATTTATTGGTAATGCCGTTATTGG 396

RESULT 15  
AA488304  
LOCUS

DEFINITION AA488304  
5' similar to gb:M21121 r-CELL SPECIFIC RANTES PROTEIN PRECURSOR (HUMAN); mRNA sequence.  
ACCESSION AA488304  
376 bp mRNA linear EST 05-MAR-1998  
aes30a04.r1 Gessler Wilms tumor Homo sapiens cDNA clone IMAGE:897294

Search completed: April 14, 2003, 22:53:33  
Job time : 1835 secs

VERSION AA488304.1 GI:2215735  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 376)  
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,  
Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin,  
J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,  
White, Y., Wylie, T., Waterston, R., and Wilson, R.  
WashU-NCI human EST Project  
Unpublished (1997)  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 1326 Std Error: 0.00  
High quality sequence stop: 315.

#### FEATURES

Location/Qualifiers  
1..376  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:897294"  
/clone\_lib="Gessler Wilms tumor"  
/sex="pooled (6)"  
/lab\_host="DH10B"  
/note="Vector: pSPORT1; Site\_1: Sali; Site\_2: NotI; RNA  
was prepared from a pool of 6 anonymous Wilms' tumor RNAs.  
RNA was prepared by acid-phenol, followed by one round of  
oligo dt selection. cDNA library preparation was with  
the BRL/Life Tech. Superscript Plasmid system. An  
oligo-dr NotI primer for first strand synthesis generated  
gcggcgccct(n at the 3' end of the clones. A 5' Sali  
adaptor was used with sequence 5'-gtcgaccacgcgtcg-3'.  
Resulting cDNAs were size selected (average size 2 kb),  
NotI digested, and ligated into NotI/Sali-cut pSPORT1.  
Library was constructed by Dr. Manfred Gessler."

BASE COUNT 97 a 77 c 75 g 127 t  
ORIGIN  
Query Match 24.8%; Score 330.4; DB 9; Length 376;  
Best Local Similarity 95.5%; Pred. No. 2.9e-67;  
Matches 340; Conservative 0; Mismatches 16; Indels 0; Gaps 0;  
QY 370 AAATGTCCTCCATTAGCATTTTCAGATCTTTTGATGTGCACATAATGCCATTATTGGTAATGC 429  
Db 1 AAATGTCCTCCATTAGCATTTTCAGATCTTTTGATGTGCACATAATGCCATTATTGGTAATGC 60  
QY 430 CGTATTGCTGAATACAGCATAGTAAATAAATACATGTTACAGTAAATACACTTGGATTT 489  
Db 61 CGTATTGCTGAATACAGCATAGTAAATAAATACATGTTACAGTAAATACACTTGGATTT 120  
QY 490 GCTGACCTCTACCAATAGCCCTTTTGAATGACTGAAAGTGTAAACAGAGAAAGGCATG 549  
Db 121 GCTGACCTCTACCAATAGCCCTTTTGAATGACTGAAAGTGTAAACAGAGAAAGGCATG 180  
QY 550 TCTGAGAAAGAGATAGCTAATATTTTGGTACTTTATCTGAATCCAGATGCTGCTT 609  
Db 181 TCTGAGAAAGAGATAGCTAATATTTTGGTACTTTATCTGAATCCAGATGCTGCTT 240  
QY 610 CCCCTGAGGTGTTGTTTCCTCTTACGATCCTCATTAATCCCTCTGGGAGCACAGGAC 669  
Db 241 CCCCTGAGGTGTTGTTTCCTCTTACGATCCTCATTAATCCCTCTGGGAGCACAGGAC 300  
QY 670 AGTAGTAGAACTCTCCATTCCTTTGTTGTTTAAAGACAGAGACTCTGTCTC 725  
Db 301 AGTAGTAGAACTCTCCATTCCTTTGTTGTTTAAAGACAGAGACTCTGTCTC 356